Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels

Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels

Adrienne Tin[1,2,196]*, Jonathan Marten[3,196], Victoria L. Halperin Kuhns[4,196], Yong Li[5,196], Matthias Wuttke[5,196], Holger Kirsten[6,7,196], Karsten B. Sieber[8], Chengxiang Qiu[9], Mathias Gorski[10,11], Zhi Yu[1,12], Ayush Giri[13,14], Gardar Sveinbjornsson[15], Man Li[16], Audrey Y. Chu[17], Anselm Hoppmann[5], Luke J. O'Connor[18], Bram Prins[19], Teresa Nutile[20], Damia Noce[21], Masato Akiyama[22,23], Massimiliano Cocca[24], Sahar Ghasemi[25,26], Peter J. van der Most[27], Katrin Horn[6,7], Yizhe Xu[16], Christian Fuchsberger[21], Sanaz Sedaghat[28], Saima Afaq[29,30], Najaf Amin[28], Johan Ärnlöv[31,32], Stephan J. L. Bakker[33], Nisha Bansal[34,35], Daniela Baptista[36], Sven Bergmann[37,38,39], Mary L. Biggs[40,41], Ginevra Biino[42], Eric Boerwinkle[43], Erwin P. Bottinger[44], Thibaud S. Boutin[3], Marco Brumat[45], Ralph Burkhardt[7,46,47], Eric Campana[45], Archie Campbell[48], Harry Campbell[49], Robert J. Carroll[50], Eulalia Catamo[24], John C. Chambers[29,51,52,53,54], Marina Ciullo[20,55], Maria Pina Concas[24], Josef Coresh[1], Tanguy Corre[37,38,56], Daniele Cusi[57,58], Sala Cinzia Felicita[59], Martin H. de Borst[33], Alessandro De Grandi[21], Renée de Mutsert[60], Aiko P. J. de Vries[61], Graciela Delgado[62], Ayşe Demirkan[28,63], Olivier Devuyst[64], Katalin Dittrich[65,66], Kai-Uwe Eckardt[67,68], Georg Ehret[36], Karlhans Endlich[26,69], Michele K. Evans[70], Ron T. Gansevoort[33], Paolo Gasparini[24,45], Vilmantas Giedraitis[71], Christian Gieger[72,73,74], Giorgia Girotto[24,45], Martin Gögele[21], Scott D. Gordon[75], Daniel F. Gudbjartsson[15], Vilmundur Gudnason[76,77], German Chronic Kidney Disease Study[78], Toomas Haller[79], Pavel Hamet[80,81], Tamara B. Harris[82], Caroline Hayward[3], Andrew A. Hicks[21], Edith Hofer[83,84], Hilma Holm[15], Wei Huang[85,86], Nina Hutri-Kähönen[87,88], Shih-Jen Hwang[89,90], M. Arfan Ikram[28], Raychel M. Lewis[91], Erik Ingelsson[92,93,94,95], Johanna Jakobsdottir[76,96], Ingileif Jonsdottir[15], Helgi Jonsson[97,98], Peter K. Joshi[49], Navya Shilpa Josyula[99], Bettina Jung[10], Mika Kähönen[100], Yoichiro Kamatani[22,101], Masahiro Kanai[22,102], Shona M. Kerr[3], Wieland Kiess[7,65,66], Marcus E. Kleber[62], Wolfgang Koenig[103,104,105], Jaspal S. Kooner[52,53,106,107], Antje Körner[7,65,66], Peter Kovacs[108], Bernhard K. Krämer[62], Florian Kronenberg[109], Michiaki Kubo[110], Brigitte Kühnel[72], Martina La Bianca[24], Leslie A. Lange[111], Benjamin Lehne[29], Terho Lehtimäki[87], Lifelines Cohort Study[78], Jun Liu[28,112], Markus Loeffler[6,7], Ruth J. F. Loos[113,114], Leo-Pekka Lyytikäinen[87], Reedik Magi[79], Anubha Mahajan[115,116], Nicholas G. Martin[75], Winfried März[62,117,118], Deborah Mascalzoni[21], Koichi Matsuda[119], Christa Meisinger[120,121], Thomas Meitinger[104,122,123], Andres Metspalu[79], Yuri Milaneschi[124], V. A. Million Veteran Program[78], Christopher J. O'Donnell[125,126], Otis D. Wilson[127], J. Michael Gaziano[128], Pashupati P. Mishra[87], Karen L. Mohlke[129], Nina Mononen[87], Grant W. Montgomery[130], Dennis O. Mook-Kanamori[60,131], Martina Müller-Nurasyid[104,132,133,134], Girish N. Nadkarni[113,135], Mike A. Nalls[136,137], Matthias Nauck[26,138], Kjell Nikus[139,140], Boting Ning[141], Ilja M. Nolte[27], Raymond Noordam[142], Jeffrey OConnell[143], Isleifur Olafsson[144], Sandosh Padmanabhan[145], Brenda W. J. H. Penninx[124], Thomas Perls[146], Annette Peters[73,74,104], Mario Pirastu[147], Nicola Pirastu[49], Giorgio Pistis[148], Ozren Polasek[149,150], Belen Ponte[151], David J. Porteous[48,152], Tanja Poulain[7], Michael H. Preuss[113], Ton J. Rabelink[61,153], Laura M. Raffield[129], Olli T. Raitakari[154,155,156], Rainer Rettig[157], Myriam Rheinberger[10], Kenneth M. Rice[41], Federica Rizzi[158,159], Antonietta Robino[24], Igor Rudan[49], Alena Krajcoviechova[160,161], Renata Cifkova[160,162], Rico Rueedi[37,38], Daniela Ruggiero[20,55], Kathleen A. Ryan[163], Yasaman Saba[164], Erika Salvi[158,165], Helena Schmidt[166], Reinhold Schmidt[83], Christian M. Shaffer[50], Albert V. Smith[77], Blair H. Smith[167], Cassandra N. Spracklen[129], Konstantin Strauch[132,133], Michael Stumvoll[168], Patrick Sulem[15], Salman M. Tajuddin[70], Andrej Teren[7,169], Joachim Thiery[7,46], Chris H. L. Thio[27], Unnur Thorsteinsdottir[15], Daniela Toniolo[59], Anke Tönjes[170], Johanne Tremblay[80,171], André G. Uitterlinden[172], Simona Vaccargiu[147], Pim van der Harst[173,174,175], Cornelia M. van Duijn[28,112,176], Niek Verweij[173,177], Uwe Völker[26,178], Peter Vollenweider[179], Gerard Waeber[179], Melanie Waldenberger[72,73,104], John B. Whitfield[75], Sarah H. Wild[180], James F. Wilson[3,49], Qiong Yang[141], Weihua

Zhang[29,52], Alan B. Zonderman[70], Murielle Bochud[181], James G. Wilson[182], Sarah A. Pendergrass[183], Kevin Ho[184,185], Afshin Parsa[186,187], Peter P. Pramstaller[21], Bruce M. Psaty[188,189], Carsten A. Böger[10,190], Harold Snieder[27], Adam S. Butterworth[191], Yukinori Okada[192,193], Todd L. Edwards[194,195], Kari Stefansson[15], Katalin Susztak[9], Markus Scholz[6,7], Iris M. Heid[11], Adriana M. Hung[127,195,197], Alexander Teumer[25,26,197], Cristian Pattaro[21,197], Owen M. Woodward[4,197], Veronique Vitart[3,197] & Anna Köttgen[1,5,197]*

1 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.
2 Welch Centre for Prevention, Epidemiology and Clinical Research, Baltimore, MD, USA.
3 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
4 Department of Physiology, University of Maryland School of Medicine, Baltimore, MD, USA.
5 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany.
6 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany.
7 LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany.
8 Target Sciences - Genetics, GlaxoSmithKline, Collegeville, PA, USA.
9 Smilow Center for Translational Research, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
10 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany.
11 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany.
12 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.
13 Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA.
14 Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA.
15 deCODE Genetics, Amgen Inc., Reykjavik, Iceland.
16 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, UT, USA.
17 Genetics, Merck & Co., Inc., Kenilworth, NJ, USA.
18 Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
19 Strangeways Research Laboratory, University of Cambridge, Cambridge, UK.
20 Institute of Genetics and Biophysics Adriano Buzzati-Traverso - CNR, Naples, Italy.
21 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy.
22 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan.
23 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.
24 Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy.
25 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany.
26 DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany.
27 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
28 Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.

29 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK.
30 Institute of Public Health & Social Sciences, Khyber Medical University, Peshawar, Pakistan.
31 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden.
32 School of Health and Social Studies, Dalarna University, Falun, Sweden.
33 Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
34 Division of Nephrology, University of Washington, Seattle, WA, USA.
35 Kidney Research Institute, University of Washington, Seattle, WA, USA.
36 Cardiology, Geneva University Hospitals, Geneva, Switzerland.
37 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland.
38 Swiss Institute of Bioinformatics, Lausanne, Switzerland.
39 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa.
40 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA.
41 Department of Biostatistics, University of Washington, Seattle, WA, USA.
42 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy.
43 Human Genetics Centre, University of Texas Health Science Centre, Houston, Texas, USA.
44 Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
45 University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy.
46 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany.
47 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany.
48 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
49 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK.
50 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
51 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore.
52 Department of Cardiology, Ealing Hospital, Middlesex, UK.
53 Imperial College Healthcare NHS Trust, Imperial College London, London, UK.
54 MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK.
55 IRCCS Neuromed, Pozzilli, Italy.
56 Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland.
57 Institute of Biomedical Technologies, Italy National Research Council, Segrate (Milano), Italy.
58 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy.
59 San Raffaele Research Institute, Milano, Italy.
60 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
61 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands.
62 5th Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany.
63 Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands.

64 Institute of Physiology, University of Zurich, Zurich, Switzerland.
65 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany.
66 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany.
67 Department of Nephrology and Medical Intensive Care, Charité – Universitätsmedizin Berlin, Germany.
68 Department of Nephrology and Hypertension, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany.
69 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany.
70 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore, MD, USA.
71 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden.
72 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.
73 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.
74 German Center for Diabetes Research (DZD), Neuherberg, Germany.
75 QIMR Berghofer Medical Research Institute, Brisbane, Australia.
76 Icelandic Heart Association, Kopavogur, Iceland.
77 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.
78 A list of members and affiliations appears in the **Supplementary Note**.
79 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia.
80 Montreal University Hospital Research Centre, CHUM, Montreal, QC, Canada.
81 Medpharmgene, Montreal, QC, Canada.
82 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, MD, USA.
83 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria.
84 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria.
85 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China.
86 Shanghai Industrial Technology Institute, Shanghai, China.
87 Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
88 Department of Pediatrics, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
89 NHLBIs Framingham Heart Study, Framingham, MA, USA.
90 The Centre for Population Studies, NHLBI, Framingham, MA, USA.
91 Department of Physiology, University of Maryland School of Medicine
92 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA.
93 Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA.
94 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden.
95 Stanford Diabetes Research Center, Stanford University, Stanford, CA, USA.

96 The Centre of Public Health Sciences, University of Iceland, Reykjavik, Iceland.
97 Landspitalinn University Hospital, Iceland.
98 University of Iceland, Iceland.
99 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, MD, USA.
100 Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
101 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.
102 Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
103 Deutsches Herzzentrum München, Technische Universität München, Munich, Germany.
104 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany.
105 Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany.
106 MRC-PHE Centre for Environment and Health, 323 School of Public Health, Imperial College London, London, UK.
107 National Heart and Lung Institute, Imperial College London, London, UK.
108 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany.
109 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria.
110 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan.
111 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, USA.
112 Nuffield Department of Population Health, Oxford University, Oxford, UK.
113 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
114 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
115 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
116 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK.
117 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany.
118 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria.
119 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.
120 Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany.
121 Chair of Epidemiology Ludwig- Maximilians-Universität München at UNIKA-T Augsburg, Augsburg, Germany.
122 Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.
123 Institute of Human Genetics, Technische Universität München, Munich, Germany.
124 Department of Psychiatry, Amsterdam Neuroscience and Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, The Netherlands.
125 VA Boston Healthcare System, Boston, MA, USA.
126 Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

127 Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville, TN, USA.
128 Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA.
129 Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.
130 University of Queensland, St Lucia, Australia
131 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The Netherlands.
132 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.
133 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany.
134 Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany.
135 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
136 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA.
137 Data Tecnica International, Glen Echo, MD, USA.
138 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany.
139 Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland.
140 Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
141 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.
142 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands.
143 University of Maryland School of Medicine, Baltimore, MD, USA.
144 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland.
145 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK.
146 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA.
147 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti (Sassari), Italy.
148 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland.
149 Faculty of Medicine, University of Split, Split, Croatia.
150 Gen-info Ltd, Zagreb, Croatia.
151 Nephrology Service, Department of Specialties in Internal Medicine, University Hospitals of Geneva, Geneva, Switzerland.
152 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK.
153 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, The Netherlands.
154 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
155 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
156 Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland.
157 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany.
158 Department of Health Sciences, University of Milan, Milano, Italy.

159 ePhood Scientific Unit, ePhood SRL, Milano, Italy.
160 Centre for Cardiovascular Prevention, First Faculty of Medicine, Department of Medicine, Charles University in Prague, Prague, Czech Republic.
161 Thomayer Hospital, Prague, Czech Republic.
162 Department of Preventive Cardiology, Thomayer Hospital, Prague, Czech Republic.
163 Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA.
164 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria.
165 Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy.
166 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria.
167 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK.
168 Division of Endocrinology, Nephrology and Rheumatology, University of Leipzig, Leipzig, Germany.
169 Heart Centre Leipzig, Leipzig, Germany.
170 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany.
171 CRCHUM, Montreal, QC, Canada.
172 Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.
173 Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
174 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
175 Durrer Centre for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, The Netherlands.
176 Leiden Academic Centre for Drug Research, Leiden University, Leiden, the Netherlands.
177 Genomics plc, Oxford, UK.
178 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany.
179 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland.
180 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK.
181 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland.
182 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson, MS, USA.
183 Geisinger Research, Biomedical and Translational Informatics Institute, Danville, PA, USA.
184 Kidney Health Research Institute (KHRI), Geisinger, Danville, PA, USA.
185 Department of Nephrology, Geisinger, Danville, PA, USA.
186 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA.
187 Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA.
188 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle, WA, USA.
189 Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA.
190 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany.
191 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

192 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan.
193 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan.
194 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, TN, USA.
195 Department of Veterans Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA.
196 These authors contributed equally to this work.
197 These authors jointly directed this project.
*E-mail: atin1@jhu.edu, anna.koettgen@uniklinik-freiburg.de

Elevated serum urate levels cause gout and correlate with cardio-metabolic diseases via poorly understood mechanisms. We performed a trans-ethnic genome-wide association study of serum urate among 457,690 individuals, identifying 183 loci (147 novel) that improve prediction of gout in an independent cohort of 334,880 individuals. Serum urate showed significant genetic correlations with many cardio-metabolic traits, with genetic causality analyses supporting a substantial role for pleiotropy. Enrichment analysis, fine-mapping of urate-associated loci, and co-localization with gene expression in 47 tissues implicated kidney and liver as main target organs and prioritized potentially causal genes and variants, including the transcriptional master regulators in liver and kidney, *HNF1A* and *HNF4A*. Experimental validation showed that *HNF4A* trans-activated the promoter of the major urate transporter *ABCG2* in kidney cells, and that HNF4A p.Thr139Ile is a functional variant. Transcriptional co-regulation within and across organs may be a general mechanism underlying the observed pleiotropy between urate and cardio-metabolic traits.

Serum urate levels reflect a balance between uric acid production and its renal and intestinal excretion. Elevated serum urate levels define hyperuricemia, which is associated with metabolic, cardiovascular and kidney-related conditions. Hyperuricemia can cause kidney stones and gout, the most common inflammatory arthritis[1,2]. Gout attacks are a highly painful response to the deposition of urate crystals, and are a significant cause of morbidity and related health care costs[3]. Although gout has become a major public health issue, it is undertreated due to low awareness, poor patient adherence[4], and inappropriate prescription practices of the most commonly used drug, allopurinol[5]. A better understanding of the mechanisms controlling serum urate may help to develop novel medications for gout treatment and prevention and provide insights into regulatory mechanisms shared between urate and cardio-metabolic traits.

Heritability of serum urate varies between 30% and 60%[6-11]. Candidate gene and genome-wide associations studies (GWAS) have identified three genes as major determinants of urate levels: *SLC2A9*, *ABCG2*, and *SLC22A12*[7,12-18]. While *SLC2A9* and *ABCG2* harbor common variants of relatively large effect[19], *SLC22A12* contains many rare or low-frequency variants[20].

The largest GWAS meta-analyses performed to date identified 28 loci among European ancestry (EA)[21] and 27 among Japanese individuals[22]. Many genes in the associated loci encode renal and intestinal urate transporters or their regulators, while others are relevant to glucose and lipid metabolism, functions of the liver, where uric acid is generated. With increased public availability of large annotation and gene expression datasets[23,24], fine-mapping associated loci to prioritize target tissues, pathways, and potentially causal genes and variants has become possible.

Here, we perform a trans-ethnic meta-analysis of GWAS of serum urate among 457,690 individuals and identify 183 associated loci that improve gout risk prediction in an independent sample of 334,880 UK Biobank (UKBB) participants. We evaluate the genetic correlation of serum urate with hundreds of cardio-metabolic traits and diseases, and use a recently developed latent causal variable model to examine the contribution of causality versus pleiotropy. We prioritize target variants, genes, tissues and pathways that contribute to the complex regulation of urate levels through comprehensive data integration. Lastly, we conduct proof-of-principle experimental studies showing that HNF4A, a transcriptional master regulator in liver and kidney proximal tubule, can regulate transcription of the major urate transporter *ABCG2* in kidney cells and that the fine-mapped *HNF4A* variant p.Thr139Ile is functional. Transcriptional co-regulation of processes linked to energy metabolism within and across organs may underlie the pleiotropy observed between urate levels and numerous cardio-metabolic traits.

Results

Trans-ethnic meta-analysis identifies 183 urate-associated loci

Trans-ethnic meta-analyses were conducted to maximize the sample size for locus discovery, and EA-specific analyses were used where population-specific linkage disequilibrium (LD) was required to characterize loci (**Supplementary Fig. 1**). The primary trans-ethnic meta-analysis included 457,690 individuals (EA, *n* = 288,649; East Asian ancestry (EAS), *n* = 125,725; African

Americans (AA), *n* = 33,671; South Asian ancestry (SA), *n* = 9,037; and Hispanics (HIS), *n* = 608) from 74 studies. Mean urate levels ranged from 4.2 to 7.2 mg/dl (**Supplementary Table 1**). GWAS were performed based on genotypes imputed using the 1000 Genomes Project or Haplotype Reference Consortium reference panels (Methods and **Supplementary Table 2**). Results were combined through inverse-variance weighted fixed effect meta-analysis after central study-specific quality control. There was no evidence of inflation due to unmodeled population structure (LD score regression intercept = 1.01; genomic inflation factor λ_{GC} = 1.04). Post-meta-analysis variant filtering left 8,249,849 high-quality SNPs for downstream analyses (Methods).

We identified 183 loci that contained at least one genome-wide significant SNP ($P \leq 5 \times 10^{-8}$, **Fig. 1** and **Supplementary Table 3**). Of these, 36 contained an index SNP reported in previous GWAS of serum urate[13,15,17,18,21,22,25,26], and 147 were considered novel (**Fig. 1**). Allelic effects on serum urate ranged from 0.28 to 0.017 mg/dl (mean 0.038 mg/dl, standard deviation (SD) 0.033). Regional association plots are shown in the **Supplementary Data Set**.

The index SNPs at all 183 loci explained 7.7% of the serum urate variance (Methods), compared to 5.3% explained by index SNPs previously reported from GWAS in EA populations[21]. In a large participating general population-based pedigree study, the 183 index SNPs explained 17% of serum urate genetic heritability (h^2 = 37%, 95% credible interval: 29%, 45%), with 5% attributed to the index SNPs at *SLC2A9*, *ABCG2* and *SLC22A12* (**Supplementary Fig. 2** and Methods).

Characterization of ancestry-related heterogeneity

For the 183 index SNPs, we observed no evidence of systematic between-study heterogeneity (median I^2 = 2%, interquartile range 0-14%; **Supplementary Table 3**). Fourteen index SNPs showed significant evidence of ancestry-associated heterogeneity ($P_{anc\text{-}het} < 2.7 \times 10^{-4}$ = 0.05/183) when tested using meta-regression (**Supplementary Fig. 3** and Methods), consistent with their higher measures of between-study heterogeneity ($I^2 > 25\%$, **Fig. 1** and **Supplementary Table 3**). The most significant ancestry-associated heterogeneity was observed

for rs3775947 at *SLC2A9* ($P_{anc\text{-}het}$ = 1.5 x 10^{-127}, allelic effect 0.34 (EA), 0.26 (AA), 0.17 (EAS), 0.41 (HIS), and 0.21 (SA) mg/dl), consistent with previous reports of population heterogeneity at this locus[27]. Nine genome-wide significant loci identified through meta-regression did not overlap with the 183 loci, including *SLC2A2* and *KCNQ1* that were genome-wide significant in EAS (**Supplementary Table 4**). Ancestry-specific meta-analyses of EA, AA, EAS and SA are summarized in **Supplementary Tables 5-8**, respectively, and in the **Supplementary Note**.

Sex-stratified meta-analyses of serum urate GWAS

Mean serum urate levels and gout risk are higher in men than in women[28]. We therefore tested whether the 183 urate-associated index SNPs showed sex-specific differences. Six SNPs showed significant effect differences (P_{diff} < 2.7 x 10^{-4} = 0.05/183), at *SLC2A9*, *ABCG2*, *CAPN1*, *GCKR*, *IDH2*, and *SLC22A12* (**Supplementary Table 9**). The genome-wide test for differences in genetic effects on urate levels between men and women identified only SNPs at *SLC2A9* and *ABCG2* (P_{diff} < 5 x 10^{-8}, Methods and **Supplementary Fig. 4**), consistent with previous reports[7,14,15,21], and several suggestive loci (P_{diff} < 1 x 10^{-5}, **Supplementary Table 10**).

Urate index SNPs are associated with gout

We next assessed the association of the 183 trans-ethnic urate index SNPs with gout in a trans-ethnic meta-analysis of 20 studies comprising 763,813 participants with 13,179 gout cases (Methods, **Fig. 1** and **Supplementary Table 1**). Consistent with the causal role of hyperuricemia in gout, genetic effects were highly correlated (Spearman correlation coefficient 0.87, **Supplementary Fig. 5a;** 0.82 for SNPs with urate association *P*-values between 5 x 10^{-8} and 1 x 10^{-8}). Fifty-five SNPs were significantly associated with gout (*P* < 2.7 x 10^{-4} = 0.05/183). In agreement with previous findings[29], the largest odds ratio (OR) for gout was observed at *ABCG2* (OR 2.04, 95% confidence interval (CI) 1.96-2.12, *P* = 7.7 x 10^{-299}). Genetic effects were generally larger among index SNPs with lower minor allele frequency (MAF), with the exception of a few common large-effect SNPs in known major urate loci *SLC2A9*, *ABCG2*, and *SLC22A12*[30] (**Supplementary Fig. 5b**).

A genetic risk score for urate improves gout risk prediction

We evaluated whether a weighted urate genetic risk score (GRS) improved gout risk prediction when added to demographic information in a large, independent sample of 334,880 UKBB participants, including 4,908 gout cases (Methods). Across categories of the GRS, gout prevalence increased from 0.1% to 12.9% (**Fig. 2a** and **Supplementary Table 11**). Compared to the most common GRS category, the age- and sex-adjusted OR of gout ranged from 0.09 (95% CI 0.02-0.37, $P = 7.8 \times 10^{-4}$) in the lowest to 13.6 (95% CI 7.2-25.7, $P = 1.4 \times 10^{-15}$) in the highest GRS category (**Fig. 2b** and **Supplementary Table 11**). The 3.5% of individuals in the three highest GRS categories had a >3-fold increase in gout risk compared to individuals in the most common GRS category. This risk is comparable to a monogenic disease of modest effect size[31], but affects a higher proportion of the population.

We additionally constructed gout risk prediction models in the UKBB sample, which was not part of the discovery analysis of serum urate-associated variants. Gout status was regressed on the GRS alone ("genetic model"), on age and sex ("demographic model"), and on the GRS, age, and sex ("combined model") in a model development subset of 90% of the individuals to obtain precise estimates. These models were then used to predict gout status in the remaining 10%, the validation sample. The genetic model was a weaker predictor (area under the receiver operating characteristic curve (AUC) = 0.68) than the demographic model (AUC = 0.79). Addition of the GRS (combined model) significantly increased prediction accuracy (AUC = 0.84, DeLong's test $P < 2.2 \times 10^{-16}$; **Fig. 2c**) and achieved a sensitivity of 84% and specificity of 68%. Ten-fold cross-validation of the regression models provided mean AUCs of 0.67 (s.d. 0.011), 0.78 (s.d. 0.006) and 0.83 (s.d. 0.008) for the genetic, demographic and combined models, respectively (Methods). The GRS represents a life-long predisposition to higher urate levels and can be calculated at birth. Thus, the GRS may help to identify individuals with a high genetic predisposition for gout, allowing for compensatory lifestyle choices to reduce the risk of gout.

High genetic correlations of serum urate with cardio-metabolic traits

Serum urate is positively correlated with many cardio-metabolic risk factors and diseases[32]. We assessed genetic correlations between urate and 748 complex traits using cross-trait LD score regression (Methods). Serum urate levels were significantly ($P < 6.6 \times 10^{-5} = 0.05/748$) genetically correlated with 214 complex traits and diseases (**Supplementary Table 12**). The highest positive genetic correlation (r_g) was with gout ($r_g = 0.92$, $P = 3.3 \times 10^{-70}$), followed by traits representing components of the metabolic syndrome such as HOMA-IR ($r_g = 0.49$) and fasting insulin ($r_g = 0.45$, **Fig. 3**). The largest negative correlations were observed with HDL cholesterol-related measurements (r_g up to -0.46), and with estimated glomerular filtration rate (r_g = -0.38 and -0.26 for cystatin C and creatinine-based estimated glomerular filtration rate (eGFR), respectively), consistent with the known role of the kidneys in urate excretion. Overall, the genetic correlations were consistent with observational associations from epidemiological studies[32].

To examine whether these genetic correlations reflect causal relationships or pleiotropy, we applied a recently developed latent causal variable (LCV) model to estimate the genetic causality proportion (GCP) for seven commonly studied cardio-metabolic traits (Methods). As a positive control, we analyzed gout, confirming a genetically causal effect of urate on gout (GCP = 0.79; **Supplementary Table 13**), consistent with Mendelian randomization (MR) studies[33,34]. The seven cardio-metabolic traits showed a GCP range consistent with mostly or partially genetically causal effects on serum urate. The largest GCP estimates were observed for adiposity-related traits (e.g. GCP = -0.84 for waist circumference; **Supplementary Table 13**), where higher cell numbers should result in higher purine and consequently urate production. A bi-directional MR study reported a causal effect of adiposity on serum urate levels[35]. While the GCP and MR methods estimate different quantities to assess causality, the direction of effect can be compared and was consistent with a positive causal effect of obesity on serum urate. Smaller GCP estimates for HDL cholesterol levels (GCP < 0.5; **Supplementary Table 13**) on the other hand suggest the existence of a genetic process with a causal effect on both HDL cholesterol and serum urate, for example co-regulated metabolic processes in the liver. These

processes may explain a large fraction of heritability for cholesterol levels and a modest fraction for urate, a type of asymmetry expected to produce a partially genetically causal relationship consistent with the one observed. MR studies did not support a causal relationship between cholesterol levels and serum urate[36].

Enriched tissues and pathways

To identify tissues and molecular mechanisms relevant for urate metabolism and handling, and to provide potential clues to the observed genetic correlations, we investigated which tissues, cell types and systems were significantly enriched for the expression of genes mapping into urate-associated loci (Methods). Based on all SNPs with $P < 1 \times 10^{-5}$, we identified significant enrichment (false discovery rate (FDR) < 0.01) for 19 physiological systems, three tissues, and two cell types (**Supplementary Table 14**). The strongest enrichment was observed for kidney ($P = 9.5 \times 10^{-9}$) and urinary tract ($P = 9.9 \times 10^{-9}$), consistent with the kidney's prominent role in controlling urate levels. Additional significant enrichments were observed for endocrine and digestive systems, including liver, the major site of urate production. Interestingly, a novel significant enrichment was also observed in the musculoskeletal system, specifically for synovial membrane, joint capsule, and joints (**Fig. 4a**), the sites of gout attacks.

We next tested for cell-type groups with evidence for enriched heritability based on cell-type-specific functional genomic elements using stratified LD score regression (Methods). The strongest enrichment was observed for kidney (11.5-fold), followed by liver (5.39-fold; **Supplementary Table 15**).

Lastly, we tested whether any gene sets were enriched for variants associated with urate at $P < 10^{-5}$ (Methods). Significant enrichment (FDR < 0.01) was observed for 383 reconstituted gene sets (**Supplementary Table 16**). Since many of these contained overlapping groups of genes, we used affinity propagation clustering to identify 57 meta gene sets (Methods and **Supplementary Table 17**), including a prominent group of inter-correlated gene sets related to kidney and liver development, morphology and function (**Fig. 4b**). Together,

these results underscore the prominent roles of the kidney and liver in regulating serum urate levels and implicate the kidney as a major target organ for lowering serum urate.

Prioritization via fine-mapping, functional annotation, and gene expression

We established a workflow that combined fine mapping of urate-associated loci with functional annotation and a systematic evaluation of tissue-specific differential gene expression to prioritize target SNPs and genes for translational research.

Statistical fine-mapping prioritizes candidate SNPs. Statistical fine-mapping was performed starting from the 123 genome-wide significant loci identified in the EA-specific meta-analysis, because the workflow included methods that used LD estimates from an ancestry-matched reference panel (Methods)[37]. After LD-based combination into 99 larger genomic regions, stepwise model selection in each region identified 114 independent SNPs ($r^2 <$ 0.01, Methods). Overall, 87 regions contained one independent SNP, ten contained two independent SNPs, the *ABCG2* locus contained three and the *SLC2A9* locus four independent SNPs (**Supplementary Table 18**). We computed 99% credible sets representing the smallest set of SNPs which collectively account for 99% posterior probability of containing the variant(s) driving the association signal (PPA)[38]. The 99% credible sets contained a median of 16 SNPs (Q1, Q3: 6, 57), and six of them only a single SNP, mapping in or near *INSR*, *RBM8A*, *MPPED2*, *HNF4A*, *CPT1C*, and *SLC2A9* (**Supplementary Table 18**). Among 28 small credible sets (≤ 5 SNPs), several mapped in or near genes with an established role in urate handling such as *SLC2A9*, *PDZK1*, *ABCG2*, *SLC22A11*, and *SLC16A9*[20]. These credible sets contain the most supported SNPs and greatly reduce the number of candidate variants for experimental follow-up.

Credible set SNPs were annotated for their functional consequence and regulatory potential (Methods). Missense SNPs with PPA > 50% or belonging to small credible sets were identified in *ABCG2*, *UNC5CL*, *HNF1A*, *HNF4A*, *CPS1*, and *GCKR* (**Fig. 5a** and **Supplementary Table 19**). All missense SNPs except the one in *GCKR* had a CADD score > 15, supporting them as potentially deleterious. Indeed, functional effects have already been demonstrated experimentally for rs2231142 (Gln141Lys, $r^2 = 1$ to the index SNP rs74904971) in *ABCG2*,

rs742493 (p.Arg432Gly) in *UNC5CL*, and rs1260326 (p.Leu446Pro) in *GCKR* (**Table 1**). Non-exonic variants with PAA > 90% and mapping into open chromatin in enriched tissues were identified in *RBM8A*, *SLC2A9*, *INSR*, *HNF4A, PDZK1*, *NRG4*, *UNC5CL*, and *AAK1* (Methods, **Supplementary Fig. 6** and **Supplementary Table 19**). When complemented by evidence of gene expression co-localization, these SNPs may represent causal regulatory variants and highlight their potential effector genes.

We compared our fine-mapping workflow ("Wakefield"), established in previous studies[39,40], to an alternative approach implemented in FINEMAP (Methods)[41]. FINEMAP identified 152 credible sets (median of 7 SNPs). With respect to known causal variants in *ABCG2* (rs2231142), *GCKR* (rs1260326), *HNF4A* (rs1800961) and *PDZK1* (rs1967017), the Wakefield approach identified the causal variants in *ABCG2*, *GCKR*, and *HNF4A* as credible set members, whereas FINEMAP found those in *ABCG2* and *HNF4A*. A comparison of all SNPs mapping into small credible sets (≤ 5 SNPs) identified through both approaches found highly correlated posterior probabilities (Pearson correlation coefficient 0.86, **Supplementary Table 19**).

Gene prioritization via gene expression co-localization analyses. The urate association signals were next tested for co-localization with expression quantitative trait loci (eQTL) in *cis* across three kidney tissue resources and 44 GTEx tissues (Methods). High posterior probability of co-localization (H4 ≥ 0.8, Methods) supports a trait-associated variant acting through gene expression in the tissue where co-localization is identified. We identified co-localization with the expression of 13 genes in kidney (**Fig. 6**), the organ with the strongest enrichment for urate-associated variants. Whereas co-localization of some genes was only observed in kidney (*SLC17A4*, *BICC1*, *UMOD*, *GALNTL5*, *NCOA7*), others showed co-localization in several tissues (e.g., *ARL6IP5*). The direction of change in gene expression with higher urate levels could vary for the same gene across tissues. For instance, the allele associated with higher serum urate at *SLC16A9* was associated with higher gene expression in kidney, consistent with a regulatory variant in a transporter mediating the reabsorption of urate. This same allele was associated with lower gene expression in other tissues such as aorta, pointing towards tissue-specific regulatory mechanisms[42]. Details of the 13 genes with evidence for co-localization with gene expression in kidney are summarized in **Supplementary Table 20**. Significant co-localizations

across all 47 tissues (**Supplementary Fig. 7**) revealed additional insights such as co-localization of the urate association signal with *NFAT5* expression in subcutaneous adipose tissue, emphasizing its role in adipogenesis[43], or *PDZK1* expression in colon and ileum, important sites of urate excretion.

Lastly, we investigated whether any trans-ethnic index SNPs or their proxies ($r^2 > 0.8$) were reproducibly associated with gene expression in *trans* in several large eQTL studies (**Supplementary Table 21** and **Supplementary Note**). We identified inter-chromosomal associations between five index SNPs and 16 transcripts that were enriched in the term "cardiovascular disease" based on the Human Disease Ontology database (**Supplementary Note** and **Supplementary Table 22**).

HNF4A activates *ABCG2* transcription and HNF4A p.Thr139Ile is a functional variant

The gene and variant prioritization workflow was validated using the identified candidates *HNF1A* and *HNF4A*. Co-regulation of target genes by these transcriptional master regulators in kidney proximal tubule and liver could potentially explain observed genetic correlations[44].

We first tested whether *HNF1A* and *HNF4A* affect transcription of *ABCG2*, which encodes for a major human urate transporter and represented the locus with the highest gout risk in our screen. The *ABCG2* promoter region contains several predicted HNF1A and HNF4A binding sites (**Fig. 5b**). A luciferase reporter assay in the human embryonic kidney cell line HEK 293 was used to assess transactivation of the human *ABCG2* promoter by HNF4A and HNF1A proteins (Methods and **Supplementary Fig. 8a**). Co-expression of HNF4A significantly increased the *ABCG2* promoter-driven luciferase activity in a transfection dose- and HNF4A protein abundance-dependent manner (**Fig. 5c** and **Supplementary Fig. 8b**). No increase of luciferase activity occurred with the negative-control vector devoid of the *ABCG2* promoter (**Supplementary Fig. 8d,e**). Results for HNF1A indicated that the observed association with serum urate is unlikely to occur via activation of *ABCG2* in kidney cells (**Fig. 5c**), but *HNF1A* has been reported to activate transcription of *PDZK1*, which encodes a regulatory protein for several other renal urate transporters[45,46] also identified in this study.

Next, we tested the functional relevance of the prioritized p.Thr139Ile allele in *HNF4A* (NM_178849.2, isoform 1, Methods). Its location within the hinge/DNA binding domain (**Fig. 5d** and **Supplementary Fig. 8f**) supports potentially altered interactions with targeted promoter regions. The isoleucine substitution at position 139 significantly increased the transactivation of the *ABCG2* promoter as compared to the wild-type threonine (**Fig. 5e**), without altering HNF4A protein abundance (**Supplementary Fig. 8c**). Thus, HNF4A can activate *ABCG2* transcription in a kidney cell line, and HNFA4 p.Thr139Ile is a functional variant. Increased activation of the urate excretory protein ABCG2 by the allele encoding the isoleucine residue should result in lower serum urate levels, consistent with the observed negative association in our GWAS.

Discussion

This trans-ethnic GWAS meta-analysis of serum urate based on 457,690 individuals represents a four-fold increase in sample size over previous studies[21,22,47] and identified 183 urate-associated loci, 147 of which are novel. A genetic urate risk score led to significant improvements of gout risk prediction among 334,880 UKBB participants: 3.5% had a risk of gout comparable to a Mendelian disease effect size. Genetic correlation and causality analyses confirmed the causal effect of urate on gout, and were consistent with transcriptional co-regulation as a source of pleiotropy in the widespread genetic correlations between serum urate and cardio-metabolic traits. Tissue and cell type-specific enrichment analyses supported kidney and liver, the sites of urate excretion and generation, as key target tissues. Comprehensive fine-mapping and co-localization analyses with gene expression across 47 tissues delivered an extensive list of target genes and SNPs for follow-up studies, of which we experimentally confirmed *HNF4A* p.Thr139Ile as a functional allele involved in transcriptional regulation of urate homeostasis.

Major challenges of GWAS are to pinpoint causal genes and variants, and to provide actionable insights into disease-relevant mechanisms. This study developed a comprehensive resource of urate-related candidate SNPs, genes, tissues and pathways that will enable a wide range of follow-up studies. Out of the many novel and biologically plausible findings, we highlight two instances in which co-localization analyses provided new insights. First, co-

localization helped to prioritize genes in association peaks that previous GWAS could not resolve. For example, the locus at chromosome 6p22.2 contains genes encoding for four members of the SLC17 transporter family (*SLC17A1-SLC17A4*). Systematic testing of co-localization across genes and tissues identified evidence only for *SLC17A4* in kidney, with higher expression associated with higher serum urate. Previous experimental studies have implicated SLC17A4 as a urate exporter in intestine[48], and our data support its yet unappreciated role in renal urate transport. Second, co-localization with *MUC1*, *BICC1* and *UMOD* expression in kidney suggests a shared biological mechanism. Rare mutations in all three genes underlie monogenic cystic kidney diseases[49-51].

Another noteworthy finding is the significant genetic correlations with many cardio-metabolic traits, consistent with observational associations[52]. Many of these traits are influenced by liver metabolism. The estimated genetic causality proportions supported their genetic correlations to be partly driven by overlapping or co-regulated metabolic pathways and not only by a fully causal effect of e.g. cholesterol or insulin levels on urate. Likewise, significant genetic correlations with kidney-related traits such as eGFR may reflect shared regulatory processes in the kidney. The observed pleiotropic effects of many urate-associated variants could thus be the potential manifestation of co-regulation of processes that occur within and across tissues relevant to the implicated traits, a mechanism likely to be prevailing in metabolic but also other traits.

In the kidney, nuclear HNF4A is exclusively detected in the proximal tubule[53], where it has been reported to regulate the expression of SLC2A9 isoform 1[54] and PDZK1[55]. Kidney-specific deletion of *Hnf4a* in mice phenocopies Fanconi renotubular syndrome[56]. Transcriptomic analyses support HNF4A to drive a proximal tubule signature cluster of 221 co-expressed genes, including many candidate genes for urate metabolism and transport[53]. In addition to *HNF4A*, *HNF4G*, and *HNF1A*, ten genes in this cluster also map into urate-associated loci we identified (*A1CF, CUBN, LRP2, PDZK1, SERPINF2, SLC2A9, SLC16A9, SLC17A1, SLC22A12* and *SLC47A1*). In addition, our study establishes that HNF4A can trans-activate transcription of *ABCG2* in a kidney cell line, the key urate secretory transporter in gut and kidney epithelium[57]. The genetic variant encoding the p.Thr139Ile substitution is located in a region of the HNF4A protein harboring

many causative mutations for monogenic maturity onset diabetes of the young (MODY type 1)[58]. Yet, unlike the severe MODY1 missense mutations p.Arg127Trp, p.Asp126Tyr, p.Arg125Trp,[59] p.Thr139Ile has not been reported to cause MODY1. Instead, it has been reported to increase the risk of type 2 diabetes, possibly through a liver-specific loss of HNF4A phosphorylation at p.Thr139, and to associate with HDL-cholesterol levels[58,60]. These data point to additional complexities when interpreting pleiotropic effects, because there may be several tissue-specific mechanisms by which genetic variants in transcriptional regulators influence metabolic pathways and urate homeostasis.

Some limitations warrant mention. The numbers of individuals of ancestries other than European or East Asian were small, and the generalizability of the gout prediction models should be assessed in future independent studies of non-European ancestry. Focusing on SNPs present in the majority of studies emphasizes those that may be of greatest importance globally over population-specific variants. General limitations of the field include that statistical fine-mapping approaches based on meta-analysis summary statistics cannot clearly prioritize functional variants in regions of tight LD, and that they are influenced by the availability and imputation quality of SNPs in the contributing studies. Only few regulatory maps from important target tissues such as synovial membrane and kidney are available, but we were able to evaluate differential gene expression in three kidney datasets. Generating additional regulatory and expression datasets across disease states, developmental stages and additional cell types in kidney and other metabolically active organs constitutes an important future research avenue. Lastly, a large independent sample for adequately powered replication testing was unavailable and represents a future endeavor. However, high correlations between genetic effects on serum urate and gout even for SNPs with the weakest significant urate associations as well as no indication of significant heterogeneity reduce concerns about false positives.

In summary, this large-scale study generated an atlas of candidate SNPs, genes, tissues and pathways involved in urate metabolism and its shared regulation with multiple cardio-metabolic traits that will enable a wide range of follow-up studies.

Acknowledgements

We thank Daniele Di Domizio (Eurac Research) and Jochen Knaus (Freiburg University) for IT assistance, and Toby Johnson (GSK) for sharing his code and discussion on credible set fine-mapping and co-localization analysis. This research has been conducted using the UK Biobank Resource under Application Number 20272. Study-specific acknowledgements and funding sources are listed in the **Supplementary Note**.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

Author contributions

Manuscript writing group: Adrienne Tin, Jonathan Marten, Victoria L. Halperin Kuhns, Yong Li, Matthias Wuttke, Holger Kirsten, Karsten B. Sieber, Chengxiang Qiu, Mathias Gorski, Markus Scholz, Adriana M. Hung, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen

Design of this study: Adrienne Tin, Jonathan Marten, Matthias Wuttke, Mathias Gorski, Christian Fuchsberger, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen

Management of an individual contributing study: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko P. J. de Vries, Alan B. Zonderman, Alessandro De Grandi, Andres Metspalu, Andrew A. Hicks, Anke Tönjes, Anna Köttgen, Annette Peters, Antje Körner, Antonietta Robino, Archie Campbell, Belen Ponte, Bernhard K. Krämer, Bettina Jung, Brenda W. J. H. Penninx, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Gieger, Christopher J. O'Donnell, Cornelia M. van Duijn, Cristian Pattaro, Daniela Toniolo, Daniele Cusi, Deborah Mascalzoni, Eric Boerwinkle, Erik Ingelsson, Florian Kronenberg, Gardar Sveinbjornsson, Georg Ehret, Gerard Waeber, Ginevra Biino, Girish N. Nadkarni, Grant W. Montgomery, Harold Snieder, Helena Schmidt, Igor Rudan, J. Michael Gaziano, James F. Wilson, James G. Wilson, Jaspal S. Kooner, Jeffrey OConnell, Joachim Thiery, Johanne Tremblay, John B. Whitfield, John C. Chambers, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Kari Stefansson, Kevin Ho, Koichi Matsuda, Konstantin Strauch, M. Arfan Ikram, Marcus E. Kleber, Marina Ciullo, Mario Pirastu, Markus Loeffler, Markus Scholz, Martin H. de Borst, Matthias Wuttke, Michael Stumvoll, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Murielle Bochud, Myriam Rheinberger, Nicholas G. Martin, Olivier Devuyst, Olli T. Raitakari, Ozren Polasek, Paolo Gasparini, Peter P. Pramstaller, Peter Vollenweider, Pim van der Harst, Qiong Yang, Rainer Rettig, Reinhold Schmidt, Renée de Mutsert, Robert J. Carroll, Ron T. Gansevoort, Ruth J. F. Loos, Sarah A. Pendergrass, Sarah H. Wild, Stephan J. L. Bakker, Tamara B. Harris, Terho Lehtimäki, Thomas Perls, Ton J. Rabelink, Uwe Völker, Vilmantas Giedraitis, Vilmundur Gudnason, Weihua Zhang, Wieland Kiess, Winfried März, Wolfgang Koenig, Yong Li, Yuri Milaneschi

Critical review of manuscript: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko P. J. de Vries, Alan B. Zonderman, Albert V. Smith, Alexander Teumer, André G. Uitterlinden, Anke Tönjes, Anna Köttgen, Annette Peters, Anselm Hoppmann, Antje Körner, Antonietta Robino, Anubha Mahajan, Audrey Y. Chu, Ayush Giri, Bernhard K. Krämer, Bettina Jung, Boting Ning, Bram Prins, Brenda W. J. H. Penninx, Brigitte Kühnel, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christa Meisinger, Christian Fuchsberger, Christian Gieger, Christopher J. O'Donnell, Cristian Pattaro, Daniel F. Gudbjartsson, Daniela Ruggiero, Deborah Mascalzoni, Dennis O. Mook-Kanamori, Erik Ingelsson, Erwin P. Bottinger, Eulalia Catamo, Florian Kronenberg, Gardar Sveinbjornsson, Ginevra Biino, Giorgia Girotto, Girish N. Nadkarni, Graciela Delgado, Grant W. Montgomery, Harold Snieder, Harry Campbell, Helgi Jonsson, Hilma Holm, Igor Rudan, Ilja M. Nolte, Ingileif Jonsdottir, Iris M. Heid, James F. Wilson, James G. Wilson, Johanna Jakobsdottir, Johanne Tremblay, John B. Whitfield, Jonathan Marten, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Karlhans Endlich, Karsten B. Sieber, Katalin Susztak, Kenneth M. Rice, Kevin Ho, Kjell Nikus, Konstantin Strauch, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Luke J. O'Connor, Man Li, Marcus E. Kleber, Marina Ciullo, Markus Loeffler, Markus Scholz, Martin H. de Borst, Martina La Bianca, Martina Müller-Nurasyid, Mary L. Biggs, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Melanie Waldenberger, Michael H. Preuss, Michele K. Evans, Mika Kähönen, Mike A. Nalls, Myriam Rheinberger, Nicholas G. Martin, Niek Verweij, Nina Hutri-Kähöne, Nisha Bansal, Olivier Devuyst, Olli T. Raitakari, Otis D. Wilson,

Ozren Polasek, Patrick Sulem, Pavel Hamet, Peter K. Joshi, Pim van der Harst, Qiong Yang, Rainer Rettig, Ravchel M. Lewis, Raymond Noordam, Renée de Mutsert, Ruth J. F. Loos, Sahar Ghasemi, Sala Cinzia Felicita, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Sarah H. Wild, Scott D. Gordon, Shih-Jen Hwang, Shona M. Kerr, Stephan J. L. Bakker, Tamara B. Harris, Teresa Nutile, Terho Lehtimäki, Thibaud S. Boutin, Thomas Meitinger, Todd L. Edwards, Ton J. Rabelink, Unnur Thorsteinsdottir, Uwe Völker, Veronique Vitart, Wei Huang, Winfried März, Wolfgang Koenig, Yong Li, Zhi Yu

Statistical methods and analysis: Albert V. Smith, Alexander Teumer, Anna Köttgen, Anselm Hoppmann, Anubha Mahajan, Audrey Y. Chu, Ayse Demirkan, Ayush Giri, Bettina Jung, Boting Ning, Bram Prins, Brigitte Kühnel, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Chris H. L. Thio, Christian Fuchsberger, Cristian Pattaro, Damia Noce, Daniel F. Gudbjartsson, Edith Hofer, Erika Salvi, Federica Rizzi, Gardar Sveinbjornsson, Ginevra Biino, Graciela Delgado, Holger Kirsten, Ilja M. Nolte, Iris M. Heid, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Jun Liu, Karsten B. Sieber, Katalin Susztak, Kathleen A. Ryan, Katrin Horn, Kenneth M. Rice, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Man Li, Marco Brumat, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Martin Gögele, Mary L. Biggs, Masahiro Kanai, Masato Akiyama, Massimiliano Cocca, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Michael H. Preuss, Mike A. Nalls, Myriam Rheinberger, Navya Shilpa Josyula, Nicola Pirastu, Niek Verweij, Nina Mononen, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Peter K. Joshi, Pim van der Harst, Qiong Yang, Raymond Noordam, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Tanguy Corre, Teresa Nutile, Thibaud S. Boutin, Todd L. Edwards, Toomas Haller, Veronique Vitart, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yoichiro Kamatani, Yong Li, Yukinori Okada

Subject recruitment: Aiko P. J. de Vries, Alan B. Zonderman, Andrej Teren, Andres Metspalu, Anke Tönjes, Anna Köttgen, Archie Campbell, Belen Ponte, Bettina Jung, Blair H. Smith, Brenda W. J. H. Penninx, Carsten A. Böger, Christa Meisinger, Cristian Pattaro, Daniela Ruggiero, Daniele Cusi, David J. Porteous, Erwin P. Bottinger, Florian Kronenberg, Gerard Waeber, Harry Campbell, Helgi Jonsson, Igor Rudan, Isleifur Olafsson, James F. Wilson, James G. Wilson, Jaspal S. Kooner, Johan Ärnlöv, Johanne Tremblay, John B. Whitfield, John C. Chambers, Katalin Dittrich, Kjell Nikus, Koichi Matsuda, Marina Ciullo, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Myriam Rheinberger, Nicholas G. Martin, Nina Hutri-Kähöne, Olli T. Raitakari, Ozren Polasek, Patrick Sulem, Peter Vollenweider, Reinhold Schmidt, Renée de Mutsert, Ron T. Gansevoort, Saima Afaq, Sandosh Padmanabhan, Sarah A. Pendergrass, Sarah H. Wild, Simona Vaccargiu, Tanja Poulain, Terho Lehtimäki, Ton J. Rabelink, Vilmundur Gudnason, Wei Huang, Winfried März

Bioinformatics: Albert V. Smith, Anna Köttgen, Anselm Hoppmann, Audrey Y. Chu, Ayush Giri, Benjamin Lehne, Bram Prins, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian M. Shaffer, Daniela Baptista, Dennis O. Mook-Kanamori, Edith Hofer, Eric Campana, Erika Salvi, Federica Rizzi, Georg Ehret, Giorgio Pistis, Holger Kirsten, Iris M. Heid, James F. Wilson, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Karen L. Mohlke, Karsten B. Sieber, Katalin Susztak, Katrin Horn, Leo-Pekka Lyytikäinen, Man Li, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Massimiliano Cocca, Mathias Gorski, Matthias Wuttke, Michael H. Preuss, Navya Shilpa Josyula, Nicola Pirastu, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Raymond Noordam, Reedik Magi, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Scott D. Gordon, Sven Bergmann, Tanguy Corre, Teresa Nutile, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yong Li, Yuri Milaneschi, Zhi Yu

Interpretation of results: Adrienne Tin, Alexander Teumer, André G. Uitterlinden, Anna Köttgen, Ayush Giri, Bettina Jung, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian Gieger, Christopher J. O'Donnell, Cristian Pattaro, Helgi Jonsson, Holger Kirsten, Iris M. Heid, Johanne Tremblay, Jonathan Marten, Karen L. Mohlke, Karlhans Endlich, Karsten B. Sieber, Katalin Dittrich, Katalin Susztak, Katrin Horn, Kevin Ho, Luke J. O'Connor, Man Li, Markus Scholz, Mathias Gorski, Matthias Wuttke, Myriam Rheinberger, Niek Verweij, Owen M. Woodward, Pavel Hamet, Pim van der Harst, Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Veronique Vitart, Victoria L. Halperin Kuhns, Wei Huang, Wolfgang Koenig, Yizhe Xu, Yong Li

Genotyping: Alan B. Zonderman, Alexander Teumer, André G. Uitterlinden, Antje Körner, Archie Campbell, Ayse Demirkan, Blair H. Smith, Brenda W. J. H. Penninx, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Fuchsberger, Cornelia M. van Duijn, Daniela Baptista, Daniela Ruggiero, Daniela Toniolo, David J. Porteous, Dennis O. Mook-Kanamori, Erik Ingelsson, Erika Salvi, Federica Rizzi, Florian Kronenberg, Georg Ehret, Grant W. Montgomery, Harry Campbell, James F. Wilson, James G. Wilson, Jaspal S. Kooner, Johan Ärnlöv, Johanne Tremblay, John C. Chambers, Karen L. Mohlke, Leo-Pekka Lyytikäinen, Leslie A. Lange, Marcus E. Kleber, Melanie Waldenberger, Michael H. Preuss, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Mike A. Nalls, Najaf Amin, Nina Mononen, Olli T. Raitakari, Patrick Sulem, Pavel Hamet, Peter Kovacs, Pim van der Harst, Ralph Burkhardt, Ron T. Gansevoort, Salman M. Tajuddin, Sandosh Padmanabhan, Scott D. Gordon, Simona Vaccargiu, Terho Lehtimäki, Thomas Meitinger, Uwe Völker, Wei Huang, Winfried März, Wolfgang Koenig, Yuri Milaneschi

Functional study: Victoria Halperin Kuhns, Raychel Lewis, Owen M. Woodward

Competing interests

Dennis O. Mook-Kanamori works as a part-time clinical research consultant for Metabolon, Inc. Brenda W. J. H. Penninx has received research funding (unrelated to the work reported here) from Jansen Research and Boehringer Ingelheim. Karsten B. Sieber is full-time employee of GlaxoSmithKline. Gardar Sveinbjornsson, Daniel F. Gudbjartsson, Ingileif Jonsdottir, Hilma Holm, Patrick Sulem, Unnur Thorsteinsdottir, and Kari Stefansson are full time employees of deCODE genetics, Amgen Inc. Audrey Y. Chu is an employee of Merck & Co., Kenilworth NJ USA. Wolfgang Koenig received modest consultation fees for advisory board meetings from Amgen, DalCor, Kowa, Novartis, Pfizer and Sanofi, and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi. Daniele Cusi is scientific consultant for Bio4Dreams. Winfried März is employed with Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. Mike A. Nalls' participation in this project is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA and consults or has consulted for during this time for Lysosomal Therapeutics Inc, Neuron23 Inc, Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Ozren Polasek is an owner of the Gen-info Ltd, Zagreb, Croatia. Kevin Ho disclosed a research and financial relationship with Sanofi-Genzyme. Bruce M. Psaty serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Adam S. Butterworth received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. Markus Scholz consults for and received grant support from Merck Serono not related to this project. Anna Köttgen received grant support from Gruenenthal not related to this project. Other authors declare no competing interests.

References

1. Kuo, C.F., Grainge, M.J., Zhang, W. & Doherty, M. Global epidemiology of gout: prevalence, incidence and risk factors. *Nat Rev Rheumatol* **11**, 649-62 (2015).
2. Li, X. *et al.* Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ* **357**, j2376 (2017).
3. Jinno, S., Hasegawa, K., Neogi, T., Goto, T. & Dubreuil, M. Trends in Emergency Department Visits and Charges for Gout in the United States between 2006 and 2012. *J Rheumatol* **43**, 1589-92 (2016).
4. Kuo, C.F., Grainge, M.J., Mallen, C., Zhang, W. & Doherty, M. Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Ann Rheum Dis* **74**, 661-7 (2015).
5. Mikuls, T.R., Farrar, J.T., Bilker, W.B., Fernandes, S. & Saag, K.G. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). *Rheumatology (Oxford)* **44**, 1038-42 (2005).
6. Yang, Q. *et al.* Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. *Metabolism* **54**, 1435-41 (2005).
7. Vitart, V. *et al.* SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* **40**, 437-42 (2008).
8. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* **2**, e132 (2006).
9. Wang, W. *et al.* Heritability and Genome-Wide Association Analyses of Serum Uric Acid in Middle and Old-Aged Chinese Twins. *Front Endocrinol (Lausanne)* **9**, 75 (2018).
10. MacCluer, J.W. *et al.* Heritability of measures of kidney disease among Zuni Indians: the Zuni Kidney Project. *Am J Kidney Dis* **56**, 289-302 (2010).
11. Rule, A.D. *et al.* Genome-wide linkage analysis for uric acid in families enriched for hypertension. *Nephrol Dial Transplant* **24**, 2414-20 (2009).
12. Enomoto, A. *et al.* Molecular identification of a renal urate anion exchanger that regulates blood urate levels. *Nature* **417**, 447-52 (2002).
13. Li, S. *et al.* The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet* **3**, e194 (2007).
14. Doring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet* **40**, 430-6 (2008).
15. Dehghan, A. *et al.* Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* **372**, 1953-61 (2008).
16. Kolz, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet* **5**, e1000504 (2009).
17. Yang, Q. *et al.* Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ Cardiovasc Genet* **3**, 523-30 (2010).

18. Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum Mol Genet* **20**, 4056-68 (2011).
19. Woodward, O.M. *et al.* Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A* **106**, 10338-42 (2009).
20. Major, T.J., Dalbeth, N., Stahl, E.A. & Merriman, T.R. An update on the genetics of hyperuricaemia and gout. *Nat Rev Rheumatol* **14**, 341-353 (2018).
21. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).
22. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
23. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).
24. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
25. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42**, 210-5 (2010).
26. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* **44**, 904-9 (2012).
27. Merriman, T.R. Population heterogeneity in the genetic control of serum urate. *Semin Nephrol* **31**, 420-5 (2011).
28. Roddy, E. & Choi, H.K. Epidemiology of gout. *Rheum Dis Clin North Am* **40**, 155-75 (2014).
29. Perez-Ruiz, F., Sundy, J.S., Miner, J.N., Cravets, M. & Storgard, C. Lesinurad in combination with allopurinol: results of a phase 2, randomised, double-blind study in patients with gout with an inadequate response to allopurinol. *Ann Rheum Dis* **75**, 1074-80 (2016).
30. Sautin, Y.Y. & Johnson, R.J. Uric acid: the oxidant-antioxidant paradox. *Nucleosides Nucleotides Nucleic Acids* **27**, 608-19 (2008).
31. Long, W. *et al.* Identification of Key Residues for Urate Specific Transport in Human Glucose Transporter 9 (hSLC2A9). *Sci Rep* **7**, 41167 (2017).
32. Feig, D.I., Kang, D.H. & Johnson, R.J. Uric acid and cardiovascular risk. *N Engl J Med* **359**, 1811-21 (2008).
33. Keenan, T. *et al.* Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. *J Am Coll Cardiol* **67**, 407-416 (2016).
34. Jordan, D.M. *et al.* No causal effects of serum urate levels on the risk of chronic kidney disease: A Mendelian randomization study. *PLoS Med* **16**, e1002725 (2019).
35. Lyngdoh, T. *et al.* Serum uric acid and adiposity: deciphering causality using a bidirectional Mendelian randomization approach. *PLoS One* **7**, e39321 (2012).
36. White, J. *et al.* Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol* **4**, 327-36 (2016).
37. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).
38. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
39. Gaulton, K.J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* **47**, 1415-25 (2015).

40. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505-1513 (2018).
41. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).
42. Pao, S.S., Paulsen, I.T. & Saier, M.H., Jr. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1-34 (1998).
43. Asano, T. *et al.* The role of N-glycosylation of GLUT1 for glucose transport activity. *J Biol Chem* **266**, 24632-6 (1991).
44. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
45. Prestin, K. *et al.* Regulation of PDZ domain-containing 1 (PDZK1) expression by hepatocyte nuclear factor-1alpha (HNF1alpha) in human kidney. *Am J Physiol Renal Physiol* **313**, F973-F983 (2017).
46. Maher, J.M. *et al.* Alterations in transporter expression in liver, kidney, and duodenum after targeted disruption of the transcription factor HNF1alpha. *Biochem Pharmacol* **72**, 512-22 (2006).
47. Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat Genet* **43**, 1127-30 (2011).
48. Togawa, N., Miyaji, T., Izawa, S., Omote, H. & Moriyama, Y. A Na+-phosphate cotransporter homologue (SLC17A4 protein) is an intestinal organic anion exporter. *Am J Physiol Cell Physiol* **302**, C1652-60 (2012).
49. Kirby, A. *et al.* Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet* **45**, 299-303 (2013).
50. Kraus, M.R. *et al.* Two mutations in human BICC1 resulting in Wnt pathway hyperactivity associated with cystic renal dysplasia. *Hum Mutat* **33**, 86-90 (2012).
51. Hart, T.C. *et al.* Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. *J Med Genet* **39**, 882-92 (2002).
52. Sodini, S.M., Kemper, K.E., Wray, N.R. & Trzaskowski, M. Comparison of Genotypic and Phenotypic Correlations: Cheverud's Conjecture in Humans. *Genetics* **209**, 941-948 (2018).
53. Lindgren, D. *et al.* Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep* **20**, 1476-1489 (2017).
54. Prestin, K. *et al.* Transcriptional regulation of urate transportosome member SLC2A9 by nuclear receptor HNF4alpha. *Am J Physiol Renal Physiol* **307**, F1041-51 (2014).
55. Ketharnathan, S. *et al.* A non-coding genetic variant maximally associated with serum urate levels is functionally linked to HNF4A-dependent PDZK1 expression. *Hum Mol Genet* **27**, 3964-3973 (2018).
56. Marable, S.S., Chung, E., Adam, M., Potter, S.S. & Park, J.S. Hnf4a deletion in the mouse kidney phenocopies Fanconi renotubular syndrome. *JCI Insight* **3**(2018).
57. Matsuo, H. *et al.* ABCG2 dysfunction causes hyperuricemia due to both renal urate underexcretion and renal urate overload. *Sci Rep* **4**, 3755 (2014).
58. Daigo, K. *et al.* Proteomic analysis of native hepatocyte nuclear factor-4alpha (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *J Biol Chem* **286**, 674-86 (2011).
59. Chandra, V. *et al.* Multidomain integration in the structure of the HNF-4alpha nuclear receptor complex. *Nature* **495**, 394-8 (2013).

60. Zhu, Q. *et al.* T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* **46**, 567-73 (2003).

Figure Legends

Figure 1 | Trans-ethnic GWAS meta-analysis identifies 183 loci associated with serum urate. Outer ring: Dot size represents the genetic effect size of the index SNP at each labeled locus on serum urate. Blue band: $-\log_{10}$(two-sided meta-analysis *P*-value) for association with serum urate (*n* = 457,690), by chromosomal position (GRCh37 (hg19) reference build). Red line indicates genome-wide significance ($P = 5 \times 10^{-8}$). Blue gene labels indicate novel loci, gray labels loci reported in previous GWAS of serum urate. Green band: $-\log_{10}$(two-sided meta-analysis *P*-value) for association with gout (*n* = 763,813), by chromosomal position. Red line indicates genome-wide significance ($P = 5 \times 10^{-8}$). Inner band: Dots represent index SNPs with significant heterogeneity and are color-coded according to its source: green for ancestry-related heterogeneity ($P_{anc\text{-}het} < 2.7 \times 10^{-4}$ (0.05/183)), red for residual heterogeneity ($P_{res\text{-}het} < 2.7 \times 10^{-4}$), and yellow for both ($P_{anc\text{-}het}$ and $P_{res\text{-}het} < 2.7 \times 10^{-4}$). Loci are labeled with the gene closest to the index SNP. $P_{anc\text{-}het}$ and $P_{res\text{-}het}$ were generated by MR-MEGA (Methods).

Figure 2 | A genetic risk score (GRS) for serum urate improves gout risk prediction. a, Histogram of the urate GRS among 334,880 European ancestry participants of the UK Biobank. The *y*-axes show the number of individuals (left) and the prevalence of gout (right), the *x*-axis shows categories of the urate GRS. The units on the *x*-axis represent genetically predicted serum urate levels (mg/dl) compared to individuals without any urate-increasing alleles. **b**, Age- and sex-adjusted odds ratio of gout (*y*-axis) by GRS category (*x*-axis) among 334,880 European-ancestry participants of the UK Biobank, comparing each category to the most prevalent category ($4.74 < GRS \leq 5.02$) with error bars representing 95% confidence intervals; * denotes logistic regression two-sided *P*-value < 0.05, ** denotes $P < 5 \times 10^{-10}$, and *** $P < 5 \times 10^{-100}$. **c**, Comparison of the receiver operating characteristic (ROC) curves of different

prediction models of gout: genetic (GRS only; red), demographic (age + sex; green), and combined (GRS + age + sex; blue). *y*-axis: sensitivity, *x*-axis: specificity. At the optimal cut points determined by the maximum of the Youden's index, the sensitivity of the combined model was 84% and specificity was 68%.

Figure 3 | Serum urate shows widespread genetic correlations with cardio-metabolic risk factors and diseases. The Circos plot shows significant genome-wide genetic correlations between serum urate and 214 complex traits or diseases (genetic correlation $P < 6.6 \times 10^{-5}$ = 0.05/748 traits tested), with bar height proportional to the genetic correlation coefficient (r_g) estimate for each trait and coloring according to its direction (dark blue, $r_g > 0$; light blue, $r_g < 0$). Traits and diseases are labeled on the outside of the plot and grouped into nine different categories. Each category is color-coded (inner ring, inset). The greatest genetic correlation was observed with gout ($r_g = 0.92$, $P = 3.3 \times 10^{-70}$). Genetic correlations with multiple cardio-metabolic risk factors and diseases reflect their known directions from observational studies. The serum urate association statistics for estimating genetic correlations were from the European-ancestry meta-analysis (*n* = 288,649).

Figure 4 | Genes expressed in urate-associated loci are enriched in kidney tissue and pathways. a, Grouped physiological systems (*x*-axis) that were tested individually for enrichment of expression of genes in urate-associated loci among European-ancestry individuals (*n* = 288,649) using DEPICT are shown as a bar plot, with the $-\log_{10}$(enrichment *P*-value) on the *y*-axis. Significantly enriched systems are labeled and highlighted in blue (enrichment false discovery rate (FDR) < 0.01). **b**, Correlated ($r > 0.2$) meta-gene sets that were strongly enriched (enrichment FDR < 0.01) for genes mapping into urate-associated loci among European-ancestry individuals (*n* = 288,649). Thickness of the edges represents the magnitude of the correlation coefficient, node size, color and intensity represent the number of clustered gene sets, gene set origin, and enrichment *P*-value, respectively.

Figure 5 | Prioritization of p.Thr139Ile at *HNF4A* and functional study of HNF4A regulation of *ABCG2* transcription. a, Graph shows credible set size (*x*-axis) against the posterior probability of association (PPA; *y*-axis) for each of 1,453 SNPs with PPA > 1% in 114 99% credible sets. Triangles mark missense SNPs, with size proportional to their Combined Annotation Dependent Depletion (CADD) score. Blue triangles indicate missense variants mapping into small (≤ 5 SNPs) credible sets or with high PPA (≥ 50%). **b**, Predicted HNF1A or HNF4A binding sites in the promoter region of *ABCG2* using LASAGNA 2.0, the consensus affinity sequence, and the *P*-value of likely matches based on nucleotide position within a consensus transcription factor binding site (Methods). **c**, Relative luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable amount of HNF1A or HNF4A constructs (mean (line) ± s.e.m. (whiskers), *n* = 3 independent experiments, *P*-values calculated with ordinary one-way ANOVA with Tukey's multiple comparison test). **d**, Position of p.Thr139Ile (T139I) in DNA binding domain/hinge region within HNF4A homodimer structure (PDB 4IQR). **e**, Relative luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable amount of constructs (ng's of transfected DNA) of wild-type HNF4A (threonine) or isoleucine at position 139 (± s.e.m., *n* = 3 independent experiments, *P*-values calculated with ordinary one-way ANOVA with Tukey's multiple comparison test).

Figure 6 | Co-localization of urate-association signals with gene expression in *cis* in kidney tissues. Serum urate association signals identified among European ancestry individuals (*n* = 288,649) were tested for co-localization with all eQTLs where the eQTL *cis*-window overlapped (±100 kb) the index SNP. Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥ 0.80) in a kidney tissue are illustrated with the respective index SNP and transcript (*y*-axis). Co-localizations across all tissues (*x*-axis) are illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 < 0.80) are marked in gray, while the positive co-

localizations are color-coded relative to the change in expression with a color gradient as indicated in the legend.

Table 1 | Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 SNPs).

Gene	SNP	#SNPs in set	SNP PP	Consequence	CADD	DHS	Gout meta-analysis *P*-value (EA)	Brief summary of literature and gene function
ABCG2	rs2231142	4	0.41	p.Gln141Lys (NP_004818.2)	18.2	ENCODE epithelial	1.21E-290	Encodes a xenobiotic and high-capacity urate membrane transporter expressed in kidney, liver and gut. Causal variants have been reported for gout susceptibility (#138900) and the Junior Jr(a-) blood group phenotype (#614490). The locus was first identified in association with serum urate through GWAS (PMID:18834626) and confirmed in many studies since. The common causal variant Q141K has been experimentally confirmed (PMID:19506252) as a partial loss of function.
UNC5CL	rs742493	4	0.95	p.Arg432Gly (NP_775832.2) (within Death domain)	21.0	ENCODE epithelial	2.73E-01	Encodes for the death-domain-containing Unc-5 Family C-Terminal-Like membrane-bound protein. Suggested as a candidate gene for mucosal diseases, with a role in epithelial inflammation and immunity (PMID:22158417). Experiments using human HEK293 cells showed that UNC5CL can transduce pro-inflammatory programs via activation of NF-κB, with the 432Gly variant less potent to do so than the 432Arg one (PMID:22158417).
HNF1A	rs1800574	2	0.92	p.Ala98Val (NP_000536.5)	23.4		1.83E-02	Encodes a transcription factor with strong expression in liver, guts and kidney. Rare mutations cause autosomal-dominant MODY type III (#600496). Locus found in GWAS of T2D (PMID:22325160) and blood urea nitrogen (PMID:29403010). Together with HNF4-alpha, it was first recognized as master regulator of hepatocyte and islet transcription. Knockout mice show proximal tubular dysfunction (Fanconi syndrome). HNF1A enhanced promoter activity of PDZK1, URAT1, NPT4 and OAT4 in human renal proximal tubule cell-based assays (PMID:28724612), supporting a role in the coordinated expression of components of the urate "transportosome".
HNF4A	rs1800961	1	1.00	p.Thr139Ile (NP_000448.3)	24.7	ENCODE pancreas	7.43E-03	Encodes another nuclear receptor and transcription factor that controls expression of many genes, including *HNF1A* and other overlapping target genes. Rare mutations cause autosomal-dominant MODY type I (#125850) and autosomal-dominant renal Fanconi syndrome 4 (# 616026). Shown to regulate expression of SLC2A9 and other members of the urate "transportosome" in cell-based assays (PMID 25209865, PMID:30124855). The GWAS locus has been reported for multiple cardio-metabolic traits and T2D (PMID:21874001).
CPS1	rs1047891	84	0.84	p.Thr1412Asn (NP_001116105.1)	22.1		5.66E-02	Encodes mitochondrial carbamoyl phosphate synthetase I, which catalyzes the first committed step of the urea cycle by synthesizing carbamoyl phosphate from ammonia, bicarbonate, and 2 molecules of ATP. Rare mutations cause autosomal-recessive carbamoylphosphate synthetase I deficiency (#237300). In addition to hyperammonemia, this disease features increased synthesis of glutamine, a precursor of purines. Elevated uric acid excretion has been reported in patients with hyperammonemia (PMID:6771064). GWAS locus for eGFR (PMID:26831199), homocysteine (PMID:23824729), urinary glycine concentrations (PMID: 26352407).
GCKR	rs1260326	2	0.67	p.Leu446Pro (NP_001477.2)	0.1	ENCODE kidney	4.09E-41	Encodes a regulatory protein prominently expressed in the liver that inhibits glucokinase. Identified in previous GWAS of urate (PMID:23263486) and multiple other cardio-metabolic traits. The 446L protein was shown to be less activated than 446Pro by physiological concentrations of fructose-6-phosphate, leading to reduced glucokinase inhibitory ability (PMID:19643913).

Abbreviation: PP, posterior probability; DHS, DNase-I hypersensitivity site; CADD, Combined Annotation Dependent Depletion phred score; EA, European ancestry.
Gout meta-analysis *P*-values were two-sided (*n* = 763,813). Posterior probabilities were estimated from statistical fine-mapping using the Wakefield approach (Methods).

Online Methods

Phenotype definition, genotyping and imputation in participating studies

The primary study outcome was serum urate in mg/dl. The laboratory methods for measuring serum urate in each study are reported in **Supplementary Table 1**. Prevalent gout was analyzed as a secondary outcome to examine whether urate-associated SNPs conferred gout risk. Gout cases were ascertained based on self-report, intake of urate-lowering medications, or International Statistical Classification of Diseases and Related Health Problems (ICD) codes for gout (**Supplementary Table 1**). The participants of all studies provided written informed consent. Each study had its research protocol approved by the corresponding local ethics committee.

Each study performed genotyping separately and imputed the genotypes to reference panels of the Haplotype Reference Consortium (HRC) version 1.1[61], 1000 Genomes Project (1000G) phase 3 v5 ALL, or the 1000G phase 1 v3 ALL[62]. Study-specific quality filters, and software used for phasing and imputation are provided in **Supplementary Table 2** and the **Supplementary Note**. Variants were annotated using NCBI b37 (hg19).

Study-specific association analysis

Phenotype generation was standardized across studies using a common script, and study-specific association analyses followed a centrally developed analysis plan. GWAS summary statistics were checked centrally using GWAtoolbox[63] and custom scripts (**Supplementary Note**). Each study performed ancestry-specific association analysis of serum urate by generating age- and sex-adjusted residuals of serum urate and regressing the residuals on SNP dosage levels, adjusting for study-specific covariates such as study centers and genetic principal components, assuming an additive genetic model. Gout was analyzed as a binary outcome adjusting for age, sex, genetic principal components, and study-specific covariates. Software used for these regression analyses were EPACTS (*q.emmax* for family based studies and *q.linear*

otherwise; https://genome.sph.umich.edu/wiki/EPACTS), SNPTest[64], RegScan[65], RVTEST[66], PLINK 1.90[67], Probabel[68], GWAF[69], GEMMA[70], mach2qtl[71] and R. Family-based studies used methods that accounted for relatedness.

Trans-ethnic, ancestry-specific, and sex-stratified meta-analyses

GWAS results from each study were pre-filtered to retain bi-allelic SNPs with imputation quality score > 0.6 and minor allele count (MAC) > 10 before inclusion into meta-analysis. Fixed effects inverse-variance weighted meta-analysis was performed using METAL[72] with modifications to output higher precision (six decimal places). Genomic control was applied for each study. The genomic inflation factor λ_{GC}[73] was calculated to assess inflation of the test statistics. For each meta-analysis result (trans-ethnic, ancestry-specific, and sex-specific), we excluded SNPs that were present in <50% of the studies and with a total MAC < 400. For ancestry-specific meta-analysis, we additionally excluded SNPs with a heterogeneity I^2-statistic[74] > 95%. Genome-wide significance was defined as *P*-value < 5×10^{-8}. The LD score regression intercept was calculated to assess the evidence for associations driven by population structure[75]. For downstream characterization, 8,249,849 and 8,217,339 autosomal SNPs were retained in the trans-ethnic and European ancestry meta-analysis, respectively. Ancestry-specific meta-analyses were conducted for European ancestry (EA), African Americans (AA), East Asian (EAS) ancestry, and South Asian (SA) ancestry using the same methods and variant filters as the trans-ethnic meta-analysis.

Secondary meta-analyses were performed separately in men and women, using the same analytical approaches. To test for significant difference of association between males and females, we used a two-sample *t*-test:

$$t = \frac{\beta_M - \beta_F}{\sqrt{SE_M{}^2 + SE_F{}^2}}$$

where β_M and β_F were beta coefficients in males and females, respectively, and SE_M and SE_F were the standard errors among males and females, respectively.

Initial determination and annotation of genome-wide significant loci

For each meta-analysis result, the SNP with the lowest *P*-value per chromosome was selected as an initial index SNP, and along with the +/- 500 kb surrounding was defined as one 1-Mb locus. This procedure was repeated with the SNP with the lowest *P*-value not yet assigned to a locus, until no genome-wide significant SNPs outside 1-Mb loci remained. To visualize loci, the genomic region +/- 500 kb around each index SNP was plotted and can contain two index SNPs when index SNPs were > 500kb but < 1 Mb apart. An ancestry-specific locus was defined as a genome-wide significant locus in an ancestry-specific meta-analysis of which the index SNP did not map into within the ±500 kb intervals of any genome-wide significant loci in the trans-ethnic meta-analysis. Index SNPs were annotated using its position and the nearest gene based on hg19, RefSeq genes, and dbSNP147 downloaded from ftp://hgdownload.soe.ucsc.edu/mysql/hg19/ on 23 March 2017.

Proportion of phenotypic variance explained and estimated heritability

The proportion of phenotypic variance explained by index SNPs was calculated as the sum of the variance explained by each index SNP based on this formula: $\beta^2 \left(\frac{2p(1-p)}{var}\right)$, where β is the beta coefficient and p is the MAF of the SNP, and var is the phenotypic variance. For this study, we used the variance of the age- and sex-adjusted residuals of serum urate in EA participants of the ARIC study as the estimate of the phenotypic variance (variance = 1.767).

Genetic heritability of age- and sex-adjusted urate levels was estimated using the R package 'MCMCglmm'[76] in the Cooperative Health Research In South Tyrol (CHRIS) study[77], a participating EA study with 4,373 individuals split into 186 up-to-five generation pedigrees[78]. Genetic heritability was estimated overall, after accounting for the index SNPs of the three major urate loci (*SLC2A9*, *ABCG2*, and *SLC22A12*), and after accounting for the index SNPs of all genome-wide significant loci for both the trans-ethnic and EA-specific meta-analyses. Estimates were obtained by running 1,000,000 MCMC iterations (*burn in* = 500,000) based on previously

described settings[78]. The difference between the overall heritability and the heritability excluding the index SNPs represents the heritability explained by the identified loci.

Trans-ethnic meta-regression

Prior to conducting trans-ethnic meta-regression, we applied the same study-specific SNP filters as those applied to the fixed effects trans-ethnic meta-analysis (imputation quality score > 0.6 and MAC > 10). An additional filter for MAF > 0.0025 was also applied to reduce the influence of rare SNPs that passed the MAC filter in very large studies. Trans-ethnic meta-regression was conducted using the MR-MEGA software package[79], which models ancestry-associated heterogeneity in the allelic effect as a function of principal components (PCs) generated from a matrix of mean pairwise allele frequency differences between studies. Three principal components generated from a matrix of mean pairwise allele frequency differences between studies were sufficient to separate the self-reported ancestry groups. Due to software requirements, the minimum number of cohorts for each SNP had to be greater than the number of PCs plus two, resulting in the exclusion of SNPs present in five or fewer cohorts. In addition to genome-wide SNP associations with urate, MR-MEGA reports ancestry-associated ($P_{anc\text{-}het}$) and residual heterogeneity ($P_{res\text{-}het}$). Index SNPs from the fixed effects meta-analysis with $P_{anc\text{-}het} < 2.7 \times 10^{-4}$ (0.05/183) in MR-MEGA were considered to have significant ancestry-associated heterogeneity.

Effect of urate-associated index SNPs on gout and risk prediction for gout

To evaluate the association of the trans-ethnic urate-associated index SNPs with gout, we conducted a trans-ethnic meta-analysis of gout with the same study-specific filtering criteria as for the urate trans-ethnic meta-analysis.

The association between a genetic urate risk score constructed from the 114 independent serum urate-associated SNPs identified among European individuals (see fine-mapping section below) and gout was assessed in a large, independent sample from the UKBB

(Projects 19655 and 20272)[80]. We selected 334,880 unrelated individuals (pairwise kinship coefficient < 0.0313) of White British ancestry with sex chromosome euploidy and concordance of phenotypic and genotypic sex, including 4,908 with gout identified by self-report at the inclusion visit. Individuals with an ICD10 for gout (M10) in hospital admissions who did not self-report gout were excluded from the analysis. A genetic risk score (GRS) was constructed as the sum of the imputed dosage of the allele associated with higher urate levels ("risk alleles") over all SNPs, multiplied by the genetic effect of the risk allele on serum urate levels. The GRS distribution was divided into ten evenly spaced categories, and individuals assigned to a category based on their GRS. The category with the lowest GRS did not contain any gout cases and so was combined with its adjacent category. Gout status was regressed on GRS category in a logistic model, including age and sex as covariates, with the category containing the largest number of individuals (genetically predicted mean urate levels 4.74-5.02 mg/dl higher compared to individuals without any urate-increasing alleles) as the reference group.

The performance of the GRS for risk prediction of gout was first evaluated in a randomly selected model development sample comprising 90% of the participants to obtain precise estimates, and tested in a validation sample of the remaining 10%. Logistic regression was used to regress gout on the GRS alone (genetic model), age and sex (demographic model) and GRS with age and sex (combined model) in the model development sample. Each of these models was then used to predict gout status in the validation sample. Model performance was assessed by comparing predicted and true gout status using Area Under the Curve (AUC) in a Receiver Operating Characteristic (ROC) curve. A cutoff of the ROC curve to report sensitivity and specificity of a combined GRS-based diagnostic test was determined by the maximum of the Youden's index (sensitivity + specificity - 1). Ten-fold cross-validation of the models was performed by randomly dividing the UKBB sample into ten equally sized groups. Each group in turn was used as the validation sample for the estimates developed on the remaining data. The AUC the ROC curve was calculated for each of the three models for all ten validation samples, and the means and standard deviations are reported.

Genetic correlation

To assess the genetic correlation between serum urate and other traits in EA, we conducted cross-trait LD score regression[81] using LD Hub[82] with the EA-specific urate meta-analysis results as input. Genetic correlation estimates with 746 traits were obtained from LD Hub, excluding two previous serum urate GWAS results. For presentation, the 212 significantly correlated traits ($P < 6.7x10^{-5} = 0.05/746$) were grouped into 9 categories based on the trait names and labels and presented in a circos plot.

To determine whether observed genetic correlations between serum urate and cardio-metabolic traits are likely to represent causal relationships, we used the recently developed latent causal variable (LCV) method to estimate the genetic causality proportion (GCP) between serum urate and another trait[83]. Compared to MR, the LCV method produces fewer false positive results in the setting of high genetic correlation and large sample sizes, a situation applicable to our analysis[83]. The GCP describes what proportion of the genetic component of one trait also affects the other trait; a positive GCP value indicates that a proportion of the genetic component of urate affects the other trait, and vice versa for a negative GCP value. LCV produces posterior mean and standard deviation estimates of the GCP using mixed fourth moments of the bivariate effect size distribution, based on GWAS summary statistics and LD scores. When using summary statistics of cardio-metabolic traits generated from the UKBB, we assumed non-overlapping populations, and overlapping populations otherwise. We selected six unique continuous cardio-metabolic traits commonly examined in epidemiological studies with high genetic correlation with serum urate ($|r_g| > 0.35$). We additionally included gout as a positive control and creatinine-based glomerular filtration rate. EA-specific GWAS summary statistics were used as input to match the ancestry of the LD scores used with the method (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2).

Functional enrichment

To assess gene-set and tissue enrichment, we used the Data-Driven Expression Prioritized Integration for Complex Traits analysis (DEPICT) version 1 release 194[84], which performs gene

set enrichment analysis by testing whether genes in 14,461 reconstituted gene sets were enriched for urate-associated SNPs (*P*-values < 1 x 10^{-5}) from the trans-ethnic meta-analysis results. Affinity propagation clustering (APC)[85], implemented in the R package 'APCluster'[86], was applied to all urate-associated reconstituted gene sets with false discovery rates (FDR)-corrected enrichment *P*-value < 0.01 to cluster gene sets containing similar combinations of genes. More details on the methods of DEPICT and APC are provided in the **Supplementary Note**. The methods for using stratified LD score regression[81] based on cell type–specific genomic annotations to identify cell type and tissue-specific enrichments of serum urate heritability are reported in the **Supplementary Note**.

Statistical fine-mapping of genome-wide significant loci in European ancestry

Statistical fine-mapping to identify potentially causal variants was performed for the genome-wide significant loci from the EA-specific meta-analysis. LD was estimated based on 16,969,363 SNPs from 13,558 unrelated UKBB participants after quality control (**Supplementary Note**). The analyses were based on a previously described workflow[39,40,87] using GCTA (cojo-slct option) to identify independent index SNPs in each region, followed by using GCTA (cojo-cond option) to obtain conditional beta and standard errors for regions with >1 independent signal. Next, approximate Bayes factors (ABF) were calculated using the Wakefield's formula[38], as implemented in the R package 'gtx' version 2.0.1 (https://github.com/tobyjohnson/gtx). The posterior probability for a variant being the driver of the association signal was calculated as the ABF of the variant divided by the sum of the ABF in the region. The 99% credible sets of a region is derived by summing the posterior probabilities in descending order until the cumulative posterior probability was > 99%. We prioritized variants in credible sets containing ≤ 5 SNPs or SNPs with posterior probabilities > 0.5. More details on statistical fine-mapping are provided in the **Supplementary Note**.

Annotation of the variants in the credible sets

We annotated SNPs in the credible sets for their exonic effect, Combined Annotation Dependent Depletion (CADD) score, and mapping into DNaseI-hypersensitive sites (DHS) from the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics Consortium projects[88,89]. The exonic effect and CADD score were obtained using SNiPA v3.2 (March 2017)[90]. SNiPA presented the CADD score as PHRED-like transformation of the C score, which was based on CADD release v1.3 downloaded from http://cadd.gs.washington.edu/download. A CADD score of 15 is used to distinguish potentially deleterious variants from background noise in clinical genetics, and represents the median value of all non-synonymous variants in CADD v1.0[91,92]. As opposed to posterior probabilities of causing the association signal, CADD scores represent an integrative measure of predicted deleteriousness based on an ensemble of variant annotations derived by contrasting common variants that survived natural selection with simulated mutations. Based on known pathogenic variants in the ClinVar database, the performance of the CADD score had an AUC of 0.88[93].

Co-localization analysis of *cis*-eQTL and urate-associated loci

Co-localization analysis of urate-associated loci with gene expression was conducted using EA meta-analysis results, *cis*-eQTL results from micro-dissected human glomerular and tubulo-interstitial kidney portions from 187 individuals in the NEPTUNE study[94], as well as from 44 tissues in the GTEx Project version 6p release[95]. For each urate locus, we identified all transcripts and all tissue-transcript pairs with reported eQTLs within ±100 kb of each GWAS index SNP. The region for each co-localization test was defined as the eQTL *cis* window in the underlying studies[94,95]. We used the default parameters and prior definitions set in the 'coloc.fast' function from the R package 'gtx' (https://github.com/tobyjohnson/gtx), which is an adapted implementation of Giambartolomei's co-localization method[24]. Evidence for co-localization was defined as H4 ≥ 0.8, which represents the posterior probability that the association with serum urate and gene expression is due to the same underlying variant. In addition, co-localization of urate-associated loci was also performed with gene expression quantified using RNA sequencing of the healthy tissue portion of 99 kidney cortex samples from

the Cancer Genome Atlas (TCGA)[96]. First, all transcripts that shared eQTL variants with urate index SNPs within ±100 kb were extracted. Then the posterior probability of co-localization was calculated including eQTLs within the *cis*-window (±1 Mb from the transcription start site) for each gene using the R coloc package[24] with default values for the three prior probabilities. The methods for *trans*-eQTL annotation are reported in the **Supplementary Note**.

Experimental study

Promoter binding site predictions. For promoter binding site predictions, we used the JASPAR 2018 database[97,98]. The frequency matrices were downloaded for transcription factor binding sites of both vertebrate and human sequences (HNF1A: MA0046.1 and MA0046.2; HNF4A: MA0114.1 and MA0114.2). These matrices were then used to query the promoter region of ABCG2 (-1285/+362, or base pairs upstream of the transcription start site / and downstream after transcription start site)[99] by means of the LASAGNA 2.0 transcription factor binding site search tool with default parameters and a P-value cutoff of 0.01[100].

Site-directed mutagenesis. HNF1A and *HNF4A* clones were purchased from GeneCopoeia (EX-A7792-M02 and EX-Z5283-M02, respectively) and were mutagenized using the QuikChange Lightning Site Directed Mutagenesis kit (Agilent Technologies, #210518) per manufacturer's instructions using PAGE purified primers, which are reported in the **Supplementary Note**.

Luciferase assay. HEK293T cells were seeded in white-walled 96-well plates coated with Poly-L-lysine at roughly 12,500 cells per well. Cells were transfected 18 hours later with either the *ABCG2* promoter (-1285/+362) upstream of a firefly luciferase in the pGL4.14 vector (a generous gift from Douglas D. Ross, University of Maryland School of Medicine), or the pGL4.14 vector (Promega, #E699A) without promoter construct, as well as GFP expressing vector used as an internal negative control (pEGFP-C1, Clontech)[101] using X-tremeGene™ 9 DNA Transfection Reagent (Roche Diagnostics, #6365787001). Transfection cocktails were prepared

per manufacturer's specifications either with or without transcription factor using the following ratio: 0.6 μg promoter construct, 0.2, 0.1, or 0.05 μg transcription factor, and 0.05 μg GFP. When no transcription factor was used, pcDNA3.1 was substituted. Approximately 48 hours after transfection, cells were rinsed with 1x PBS, then lysed using Passive Lysis Buffer (Promega #E194A) for 15 minutes. During this incubation, GFP measurements were taken using a CLARIOstar microplate reader (BMG Labtech). Next, 30 μl of Luciferase Reagent (Promega, E297A&B) were added to each well, and the plate was incubated for an additional 20 minutes at room temperature. Finally, luciferase activity was measured using the CLARIOstar microplate reader taking the average over 6 seconds. To evaluate the significance of transactivation of the *ABCG2* promoter, we compared cells expressing transcription factors to those transfected with the empty vector (pcDAN3.1) and to evaluate TF dose responses or differences in TF variants all experimental conditions from one plate were compared using an Ordinary one-way ANOVA, accounting for multiple comparisons with a Tukey's multiple comparison test. Statistical analysis was performed using Prism 7 (GraphPad Software Inc, USA).

Western blots. Equal volumes of deoxycholate-RIPA buffer were added to wells containing desired lysates following the luciferase assay and plates were then incubated at 4 °C overnight. Equal volumes of sample + 5x SDS loading dye + 10% β-merceptoethanol were then loaded into 10% Mini-PROTEAN® TGX Stain-Free™ Precast Gels (Bio-Rad, #4568033) and run per manufacturer's specifications. Gels were then cross-linked for 45 seconds and imaged to reveal total protein load, which was used as the loading control for each lane (representative images of these protein gels are found in **Supplementary Fig. 8**). Gels were then transferred onto nitrocellulose membranes using the Trans-Blot® Turbo™ Transfer System (Bio-Rad), blocked for 2 hours at room temperature in 5% milk in TBS-T, and incubated overnight at 4 °C with primary antibody. Membranes were then washed 3 times with TBS-T, incubated at room temperature for 1 hour with Donkey anti-rabbit secondary antibody (Jackson ImmunoResearch, #111-035-144) diluted 1:5,000 in 2.5% milk in TBS-T. Membranes were then washed again and developed using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, #34577) and imaged on the ChemiDoc MP imaging system (Bio-Rad). All primary antibodies were diluted

1:1,000 in 2.5% milk in TBS-T. Antibodies used included HNF4α (Cell Signaling Technology, #3113) and HNF1α (Cell Signaling Technology, #89670). Antibodies were validated using lysates of overexpressing HEK293T cells transfected with either HNF construct, demonstrating bands at the appropriate sizes (**Supplementary Fig. 8**).

Reporting Summary. Further information on research design is available in the **Nature Research Reporting Summary** linked to this article.

Data availability

Genome-wide summary statistics for this study are shared at http://ckdgen.imbi.uni-freiburg.de and will be made publicly available through dbGaP accession number phs000930.v6.p1.

Methods-only References

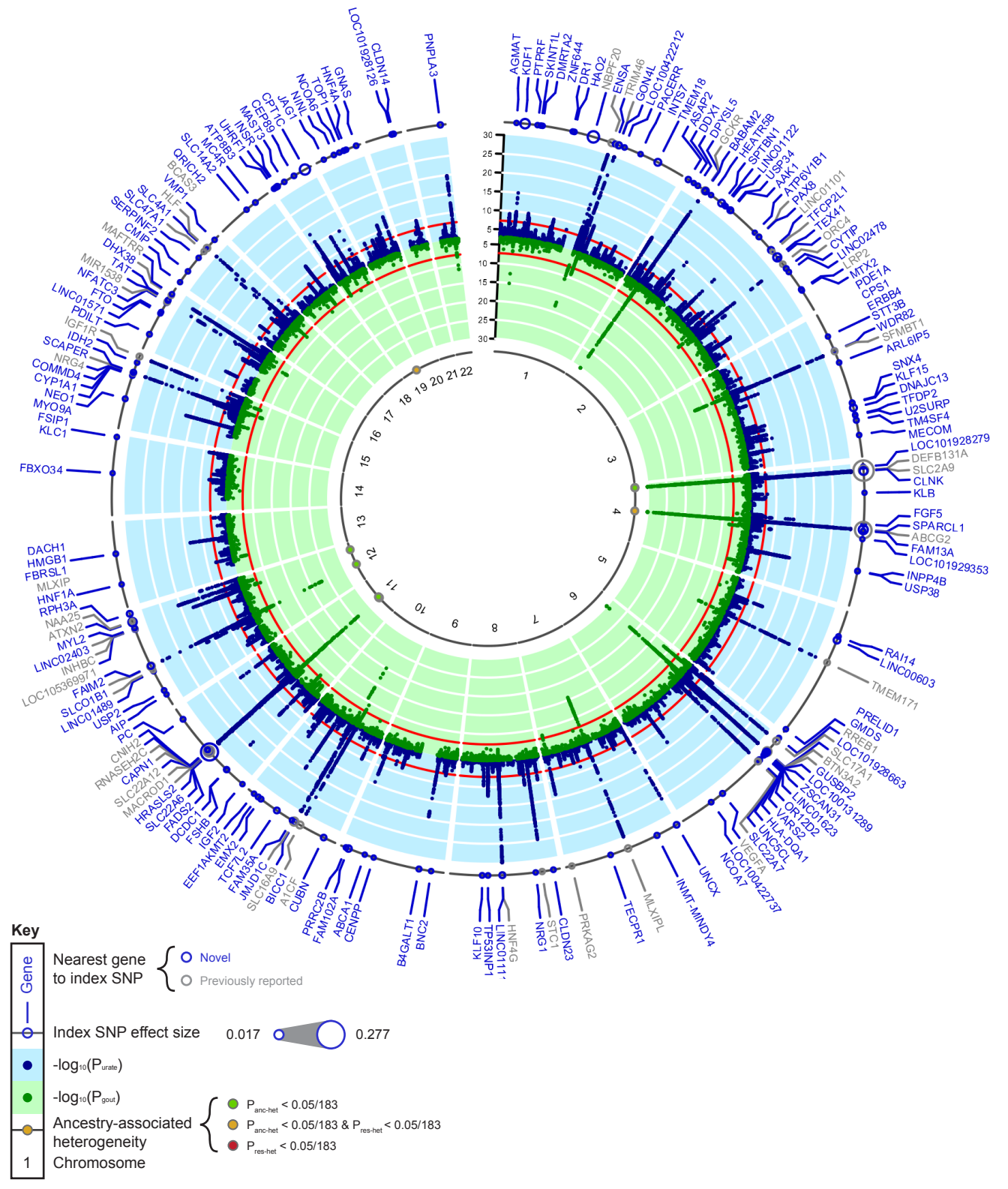
61. The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283 (2016).
62. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
63. Fuchsberger, C., Taliun, D., Pramstaller, P.P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444-5 (2012).
64. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).
65. Haller, T., Kals, M., Esko, T., Magi, R. & Fischer, K. RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. *Brief Bioinform* **16**, 39-44 (2015).
66. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-6 (2016).
67. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
68. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
69. Chen, M.H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580-1 (2010).
70. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-4 (2012).
71. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).
72. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
73. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155-66 (2001).
74. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
75. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
76. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).

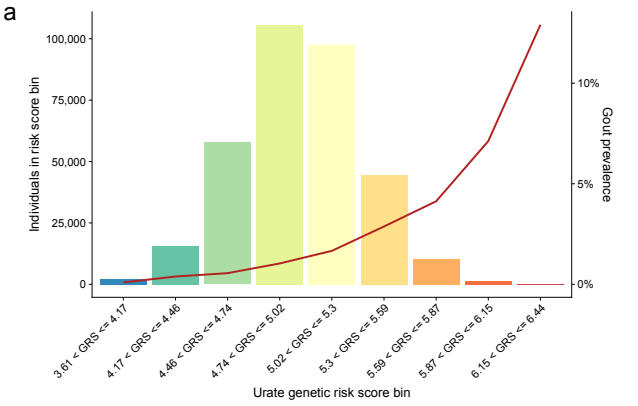
77. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).
78. Noce, D. *et al.* Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum Genet* **136**, 743-757 (2017).
79. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).
80. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
81. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
82. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
83. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* **50**, 1728-1734 (2018).
84. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
85. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6 (2007).
86. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463-4 (2011).
87. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* **51**, 957-972 (2019).
88. Sheffield, N.C. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res* **23**, 777-88 (2013).
89. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
90. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
91. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-37 (2015).
92. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
93. Li, J. *et al.* Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res* **46**, 7793-7804 (2018).
94. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* **103**, 232-244 (2018).
95. GTEx. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
96. Ko, Y.A. *et al.* Genetic-Variation-Driven Gene-Expression Changes Highlight Genes with Important Functions for Kidney Disease. *Am J Hum Genet* **100**, 940-953 (2017).
97. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D1284 (2018).
98. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4 (2004).
99. Xie, Y. *et al.* Functional cyclic AMP response element in the breast cancer resistance protein (BCRP/ABCG2) promoter modulates epidermal growth factor receptor pathway-

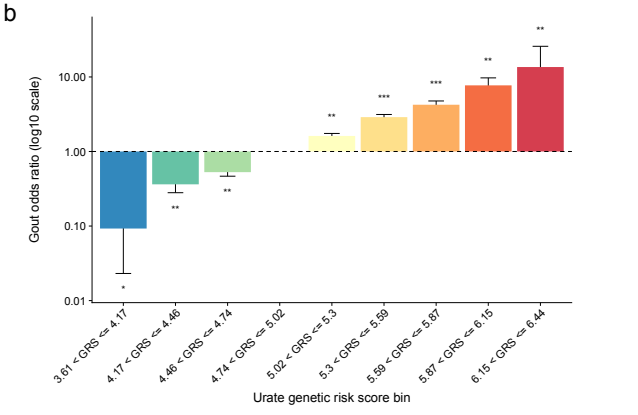
or androgen withdrawal-mediated BCRP/ABCG2 transcription in human cancer cells. *Biochim Biophys Acta* **1849**, 317-27 (2015).
100. Lee, C. & Huang, C.H. LASAGNA-Search 2.0: integrated transcription factor binding site search and visualization in a browser. *Bioinformatics* **30**, 1923-5 (2014).
101. Vesuna, F., Winnard, P., Jr. & Raman, V. Enhanced green fluorescent protein as an alternative control reporter to Renilla luciferase. *Anal Biochem* **342**, 345-7 (2005).

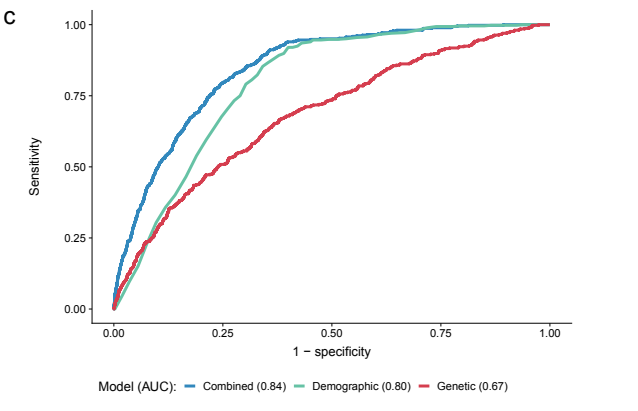
Editorial summary:

A trans-ethnic genome-wide association study of serum urate levels identifies 183 loci influencing this trait. Enrichment analyses, fine mapping and co-localization with gene expression in 47 tissues implicate kidney and liver as key target organs and prioritize potential causal genes.


Key
Gene
Nearest gene to index SNP
Novel
Previously reported
Index SNP effect size
0.017
0.277
-log10(Purate)
-log10(Pgout)
Ancestry-associated heterogeneity
Panc-het < 0.05/183
Panc-het < 0.05/183 & Pres-het < 0.05/183
Pres-het < 0.05/183
1
Chromosome

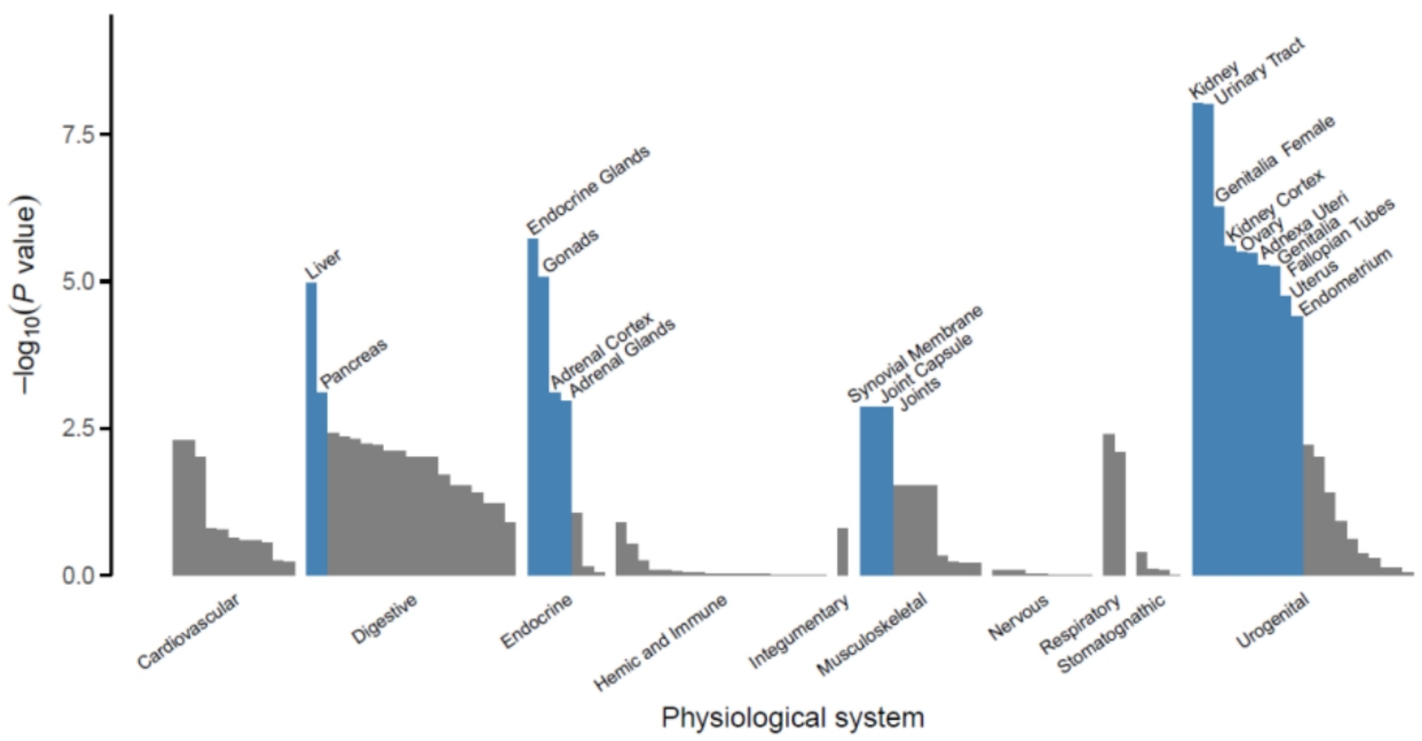

a
Individuals in risk score bin
100,000
75,000
50,000
25,000
0
Gout prevalence
10%
5%
0%
3.61 < GRS <= 4.17
4.17 < GRS <= 4.46
4.46 < GRS <= 4.74
4.74 < GRS <= 5.02
5.02 < GRS <= 5.3
5.3 < GRS <= 5.59
5.59 < GRS <= 5.87
5.87 < GRS <= 6.15
6.15 < GRS <= 6.44
Urate genetic risk score bin


b
Gout odds ratio (log10 scale)
10.00
1.00
0.10
0.01
*
**
**
**
***
***
**
**
3.61 < GRS <= 4.17
4.17 < GRS <= 4.46
4.46 < GRS <= 4.74
4.74 < GRS <= 5.02
5.02 < GRS <= 5.3
5.3 < GRS <= 5.59
5.59 < GRS <= 5.87
5.87 < GRS <= 6.15
6.15 < GRS <= 6.44
Urate genetic risk score bin

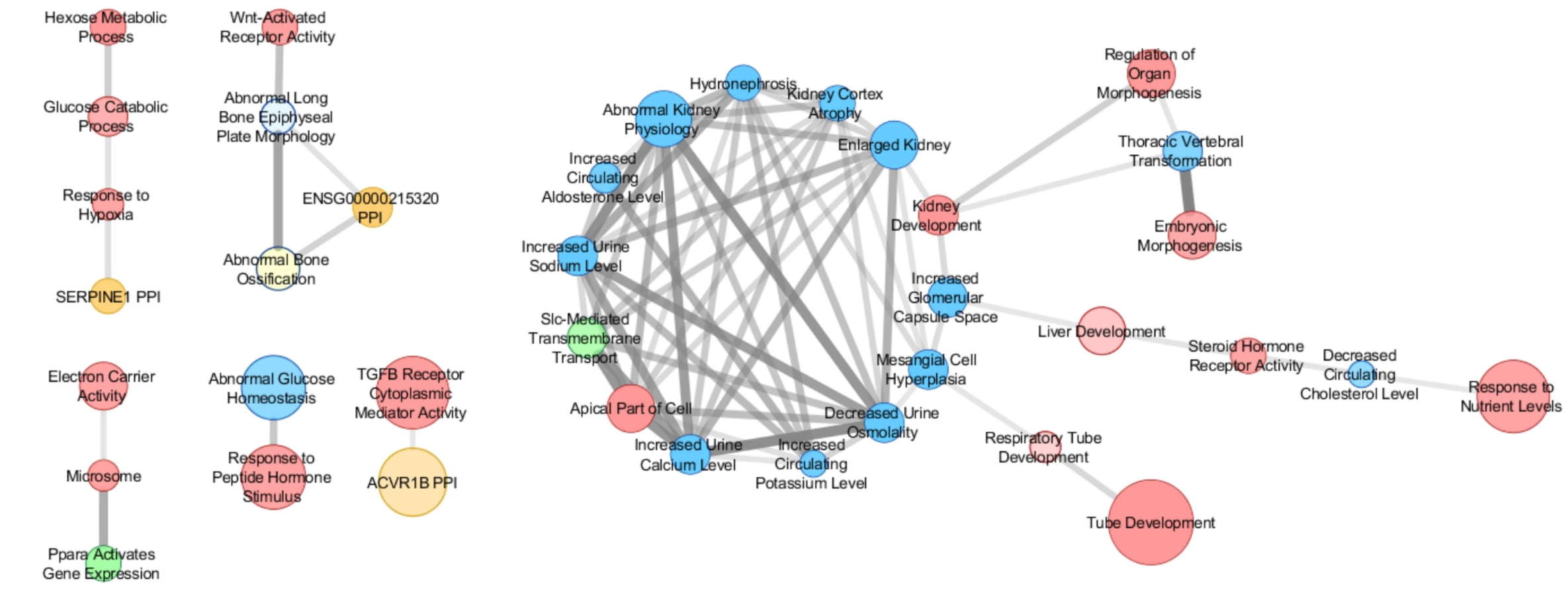

c
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − specificity
Model (AUC):
Combined (0.84)
Demographic (0.80)
Genetic (0.67)


0.75
0.5
0.25
0
Gout
Alcohol and smoking
Amino acid, protein and lipid
Anthropometric
Cardiovascular and kidney
Education and behavior
Lung function, physical activity
Medication
Sleeping and others
Self−reported gout
Why quit smoking: health precaution
Alcohol taken with meals
Weekly champagne intake
Weekly red wine intake
Maternal smoking around birth
Exposure to smoke outside home
Weekly spirits intake
Alcohol intake frequency
Previously daily cigarettes
Alcohol intake vs. 10 years ago
Pack years adult smoking
Exposure to smoke at home
Free cholesterol in large HDL
Total cholesterol in large HDL
Cholesterol esters in large HDL
Phospholipids in large HDL
Total lipids in large HDL
Total cholesterol in HDL
Large HDL particles
Mean diameter for HDL particles
HDL cholesterol
Cholesterol esters in medium VLDL
Very large VLDL particles
Total lipids in very large VLDL
Total lipids in largest VLDL
Cholesterol in medium VLDL
Free cholesterol in small VLDL
Cholesterol esters in very small VLDL
Triglycerides in large VLDL
Cholesterol esters in small VLDL
Phospholipids in small VLDL
Total lipids in small VLDL
Total lipids in large VLDL
Large VLDL particles
Free cholesterol in large VLDL
Small VLDL particles
Glycoprotein acetyls
Mean diameter for VLDL particles
Phospholipids in chylomicrons/largest VLDL
Phospholipids in very large VLDL
Total lipids in medium VLDL
Triglycerides in very large VLDL
Total cholesterol in large VLDL
Medium VLDL particles
Free cholesterol in medium VLDL
Phospholipids in large VLDL
Triglycerides in chylomicrons/largest VLDL
Self−reported high cholesterol
Isoleucine
Serum total triglycerides
Phospholipids in medium VLDL
Triglycerides in large VLDL
Triglycerides
Triglycerides in small VLDL
Triglycerides in medium VLDL
Triglycerides in small HDL
Impedance of whole body
Impedance of right arm
Impedance of left arm
Impedance of left leg
Impedance of right leg
Birth weight
Trunk predicted mass
Trunk fat−free mass
Comparative body size age 10
Whole body fat−free mass
Whole body water mass
Childhood obesity
Right arm fat−free mass
Right arm predicted mass
Right leg fat−free mass
Right leg predicted mass
Left arm fat−free mass
Left arm predicted mass
Left leg fat−free mass
Left leg predicted mass
Hip circumference − UKBB
Weight
Waist−to−hip ratio
Trunk fat percentage
Trunk fat mass
Right leg fat percentage
Right leg fat mass
Body fat percentage
Left leg fat percentage
Left leg fat mass
Body fat
Overweight
Hip circumference
Whole body fat mass
Right arm fat percentage
Left arm fat percentage
Obesity class 1
Body mass index
Left leg fat mass
Right leg fat mass
Body mass index − UKBB
Left arm fat mass
Right arm fat mass
Extreme bmi
Waist circumference − UKBB
Obesity class 3
Obesity class 2
Waist circumference
Obesity class 1
No siblings illness
Fathers age at death
No diagnosed heart problem
No mother illness
Parents age at death
Mothers age at death
Fathers age at death − UKBB
eGFR
MM−DM
eGFR non−DM
No father illness
Urine sodium
Diabetic disease father
Heart disease father
Heart disease mother
Fasting glucose main effect
Coronary artery disease
Hypertension mother
Systolic blood pressure
Hypertension father
Diastolic blood pressure
Self−reported heart attack/MI
Diagnosed heart attack
Heart disease siblings
Angina pectoris
HOMA−B
Diabetes mother
Chronic ischaemic heart disease
Hypertension siblings
Diagnosed hypertension
Self−reported hypertension
Hypertension siblings
Diagnosed angina
Self−reported angina
Diabetes siblings
Type 2 diabetes
Diagnosed diabetes
Self−reported diabetes
Leptin_not_adjBMI
Chronic kidney disease
Diabetes related eye disease
Self−reported type 2 diabetes
Fasting insulin main effect
HOMA−IR
Age completed education
Years of schooling 2013
Education: A levels qualifications
Years of schooling (proxy)
Years of schooling 2016
Education: college degree
Education: no qualifications
Time spent watching TV
Maximum workload during fitness test
Usual walking pace
Non−work transport: cycle
Work transport: cycle
Physical activity: strenuous sports
Stair climbing frequency
Physical activity: walking
Physical activity: heavy DIY
Best FVC
FVC − UKBB
FEV1
Best FEV1
Financial difficulties
Townsend deprivation index
Frequency of tiredness/lethargy
Currently unable to work
Knee pain
Wheeze or whistling in chest
Physical activity: none
Working week length main job
Serious illness, injury or assault
Overall health rating
Basal metabolic rate
Health satisfaction
Leg pain on walking
Frequency travelling from home to job
Chest pain or discomfort walking
Shortness of breath walking
Number of self−reported illnesses
No medication in men
No medication in women
No pain relief medication
Paracetamol medication
Other prescription medications
Number of treatments/medications
Hypertensive medication in women
Hypertensive medication in men
Cholesterol medication in men
Aspirin medication
Cholesterol medication in women
Acceleration standard deviation
Relative age voice broke
Frequency of light DIY
Age at menarche − UKBB
Age at last live birth
Age of first live birth
Age at menarche
Relative age of first facial hair
Snoring
Nap during day
Schizophrenia
Left heel bone mineral density
Self−reported hypothyroidism
Pain in throat and chest
Education: CSEs
Abdominal and pelvic pain
Cholelithiasis
Genetic correlation (rg)
rg > 0
rg < 0

a



b



Gene set origin:
- Gene Ontology
- Mouse Phenotype
- PPI network
- REACTOME pathway

DEPICT P-value: colour intensity

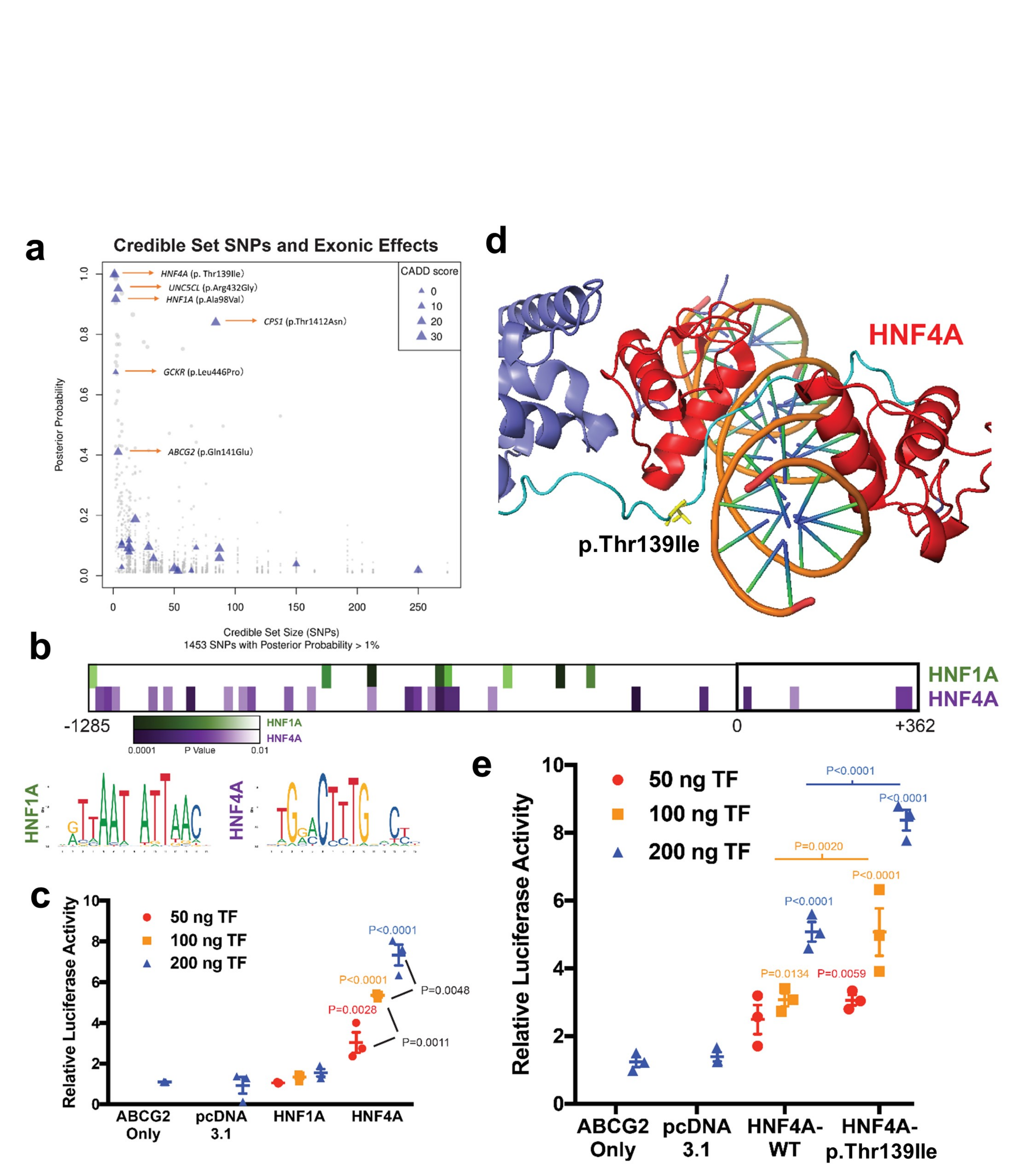
1.68×10^{-10} — 4.5×10^{-4}

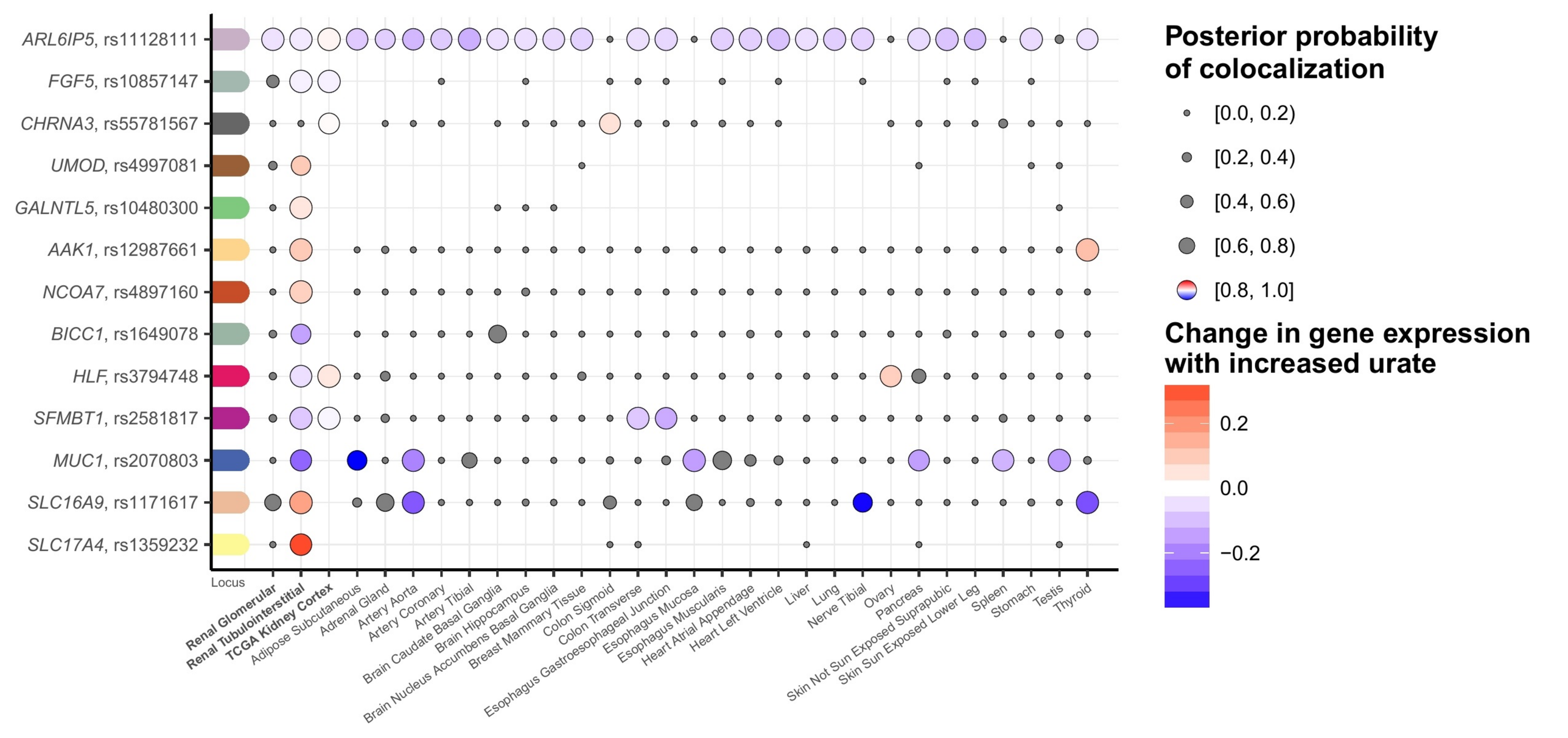
Cluster size: Node size

1 — 28

Correlation coefficient: edge

0.2 — 0.89


a
Credible Set SNPs and Exonic Effects
HNF4A (p. Thr139Ile)
UNC5CL (p.Arg432Gly)
HNF1A (p.Ala98Val)
CPS1 (p.Thr1412Asn)
GCKR (p.Leu446Pro)
ABCG2 (p.Gln141Glu)
CADD score
0
10
20
30
Posterior Probability
Credible Set Size (SNPs)
1453 SNPs with Posterior Probability > 1%
d
HNF4A
p.Thr139Ile
b
HNF1A
HNF4A
-1285
0
+362
P Value
0.0001
0.01
HNF1A
HNF4A
c
Relative Luciferase Activity
50 ng TF
100 ng TF
200 ng TF
P<0.0001
P<0.0001
P=0.0048
P=0.0028
P=0.0011
ABCG2 Only
pcDNA 3.1
HNF1A
HNF4A
e
Relative Luciferase Activity
50 ng TF
100 ng TF
200 ng TF
P<0.0001
P<0.0001
P=0.0020
P<0.0001
P<0.0001
P=0.0134
P=0.0059
ABCG2 Only
pcDNA 3.1
HNF4A-WT
HNF4A-p.Thr139Ile


ARL6IP5, rs11128111
FGF5, rs10857147
CHRNA3, rs55781567
UMOD, rs4997081
GALNTL5, rs10480300
AAK1, rs12987661
NCOA7, rs4897160
BICC1, rs1649078
HLF, rs3794748
SFMBT1, rs2581817
MUC1, rs2070803
SLC16A9, rs1171617
SLC17A4, rs1359232
Locus
Renal Glomerular
Renal Tubulointerstitial
TCGA Kidney Cortex
Adipose Subcutaneous
Adrenal Gland
Artery Aorta
Artery Coronary
Artery Tibial
Brain Caudate Basal Ganglia
Brain Hippocampus
Brain Nucleus Accumbens Basal Ganglia
Breast Mammary Tissue
Colon Sigmoid
Colon Transverse
Esophagus Gastroesophageal Junction
Esophagus Mucosa
Esophagus Muscularis
Heart Atrial Appendage
Heart Left Ventricle
Liver
Lung
Nerve Tibial
Ovary
Pancreas
Skin Not Sun Exposed Suprapubic
Skin Sun Exposed Lower Leg
Spleen
Stomach
Testis
Thyroid
Posterior probability of colocalization
[0.0, 0.2)
[0.2, 0.4)
[0.4, 0.6)
[0.6, 0.8)
[0.8, 1.0]
Change in gene expression with increased urate
0.2
0.0
−0.2